ISSN 0001 253X
Volume 52
Number 7
July/August 2000

Cu cl
Q. 18/10/
H04728-51-P035540
51

Aslib Proceedings

New information perspectives

Practice makes perfect? Vets' information seeking behaviour and information use explored
Tim Wales

The century of film: bibliographic control and legal deposit of the moving image
Stephen Pearson

The European information needs of secondary school teachers in Scotland: recent developments in the provision of information to schools and colleges
Gillian Conroy, Susan Parker and Sylvie Davies

Visions of the information future
Alison Scammell

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

Aslib Proceedings is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000–6,000 words in length. Articles should be accompanied by a 100–150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The editor is happy to provide advice on an informal basis to potential authors.

Editorial Board

Dr David Nicholas, *Head of Department, Department of Information Science, City University*; John Akeroyd, Director, *Learning Resources, South Bank University*; Sarah Blair, *Head of Publications, Aslib*; Peter Chapman, *Information Manager, NewQuest North East*; Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire*; Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw*; Harry East, *Senior Research Fellow, Department of Information Science, City University*; Mike Hyman, *Managing Director, Head Software International*; Robin Hunt, *Creative Director, AreHause*; Nat Lievesley, *Network Manager, Centre for Policy on Ageing*; Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University*; Richard Withey, *Director, Independent Digital*; Ralph Adam, *Freelance Journalist*; Professor David Ellis, *Department of Information and Library Studies, University of Wales;* Professor Ben Fouche, *Stellenbosch University*.

Editorial Team

Editor: David Nicholas
Deputy Editor: Ian Rowlands
Editorial Manager: Kay Nicholas
Editorial Assistants: Peter Williams (City University), Colette Coles (London Research Centre)

P 35,540

CONTENTS

235–246 Practice makes perfect? Vets' information seeking behaviour and information use explored
Tim Wales

247–253 The century of film: bibliographic control and legal deposit of the moving image
Stephen Pearson

254–263 The European information needs of secondary school teachers in Scotland: recent developments in the provision of information to schools and colleges
Gillian Conroy, Susan Parker and Sylvie Davies

264–269 Visions of the information future
Alison Scammell

Aslib Proceedings Vol 52, No.7, July/August 2000 – 234

Practice makes perfect? Vets' information seeking behaviour and information use explored

Tim Wales
Department of Information Science, City University, Northampton Square, London, EC1V 0HB
tim@twales.freeserve.co.uk

A random sample of UK veterinary practitioners was surveyed and interviewed on behalf of the Royal College of Veterinary Surgeons Wellcome Library to identify key issues in veterinary information use (IU) and information seeking behaviour (ISB).[1] A greater proportion of respondents used the Internet for veterinary information than used a veterinary library. However, conventional journals, textbooks and conferences were the main information sources used. Some variations in information source use by practice size and type and information type were identified. The majority of library users and non-users wanted enhanced library access via the Internet, especially to full-text journals.

Introduction

Librarianship and Information Science (LIS) research in the field of veterinary science is rare – arguably, the last UK LIS study was carried out by a student at City University in 1993, before the 'information explosion' occasioned by the Internet. This means that veterinary information services (VIS) have been effectively operating in a vacuum with regard to their users' needs, especially where electronic information is concerned.

This has particular resonance for vets in practice working 'on the front line', alone or as part of a small team, often based in rural areas, far away from VIS that are, by and large, intended (and funded) to support first and foremost the veterinary research undertaken by government or academic vets. Practising vets who need help sourcing information will ultimately turn to the Royal College of Veterinary Services Welcome Library (RCVSWL), the 'profession's library' [1].

Accordingly, this research was undertaken on behalf of RCVSWL to determine present and future service issues by examining a sample of veterinarians in practice, to discover where and by what means they looked for the information they needed and what use they made of it once found. It was intended to be a localised, microenvironmental study, which would complement a future macroenvironmental study on vets' information needs (IN).

Previous research

Pelzer and Leysen's 1991 survey of American vets can be considered the definitive veterinary ISB/IU research to date. It supported previous research by Drake and Woods which established that books were the primary information source used in critical care situations, and found that relatively little use was made of veterinary medical libraries, computer databases or extension services [2, 3]. The latter point was highlighted as having potentially negative implications for vets' awareness of current issues in preventative medicine and zoonoses. Journals were perceived as the most important source for 'keeping up-to-date', despite the fact that the information contained within would be at least one year old. Finally, although 50% of respondents had the use of a computer, hardly any used it for database searching.

---

[1] A full version of this research (including the original questionnaire) is available online at http//www.twales.freeserve.co.uk/WALES/Infosurvey_99_for_web.pdf

The same authors' research (1988 [4], updated 1998 with Wiese [5]) into the library use and ISB of a cross-section of US veterinary medical students is useful mainly for its section on student use of electronic resources. VETCD, MEDLINE and BEASTCD were the most popular databases used, illustrating the general shift from print to computer indexes in library ISB/IU since 1988. In general, American veterinary students showed themselves to be pragmatic electronic resource users, acknowledging that other resources such as colleagues, CPD courses/workshops and conferences would continue to be their main sources of information.

The few studies of UK vets that have been undertaken tend towards the enumeration and quantification of veterinary information held in practices: numbers of books held, journals subscribed to, etc., as opposed to a deeper investigation of IN, IU or ISB. However, Bawden and Valleley's 1996 [6] (fieldwork 1993 [7]) survey is a good reference for its findings on computer use in practices. Little use was made of computers for storage or retrieval of veterinary information – accounting and administrative tasks predominated. There was a general lack of awareness of electronic resources with only 10% of respondents being aware of specific information sources. One third of respondents were happy with the amount of information available and the means of obtaining it. The remaining two thirds expressed some degree of dissatisfaction, being time-pressed vets who saw themselves faced with a mountain of information. This could only be addressed by VIS ensuring quick and easy access to information.

Raw had come to a similar conclusion some years previously [8]. He drew on a larger sample (537 practices) and chose to focus on practice library issues – subscriptions, book selection criteria, budgets etc. Respondents accused publishers of selling out-of-date books (due to the long lead times involved in preparing new editions) and preferred specialist material aimed at vets in practice rather than specialist 'esoteric' research material appearing in journals. Raw pointed out that such research eventually filters down into the practice of tomorrow, but missed the fact that his respondents were requesting a selective dissemination of information (SDI) service in all but name.

In Gerrard's research on veterinary practitioners' views on the Internet, it was reported that, although the level of Internet awareness was high, most vets were unsure about the Internet and the benefits it could offer them and that this uncertainty was the primary reason for vets not using it [9]. Of his sample 46% had no interest in CPD courses aimed at increasing Internet awareness – a key issue, the author reckoned, if the Internet is to be used to its full potential by the profession.

Biggs found that 75% of veterinary surgeons in general practice in Britain used computers as a work-aid, with 90% of recently qualified vets being computer users [10]. Seventy-five per cent of vets had access to a CD-ROM facility but less than half were on the Internet. Information on diseases and veterinary products (and their manufacturers) were the main focuses of Internet searches. Computer use was found to absorb on average under 10% of a vet's time. Finally, the author judged veterinary computer use to be 'fairly basic' as statistical analysis and accounting were found to be minority uses. This contrasts with Bawden and Valleleys' findings.

Relevant ISB/IU research from other disciplines

Various medical ISB/IU studies of practitioners exist. Gravois *et al.* studied dental hygienists and suggested information evaluation and perceived barriers to information use as two worthy research areas [11]. Bowden *et al.* examined IN/ISB/IU of Texas physicians with particular attention to the availability of MEDLINE [12]. The authors found that personal collections of books and journals and consultations with colleagues were more important to physicians than MEDLINE. Lundeen *et al.* found geographic isolation, cost and inadequate technology to be the principal barriers to information use cited by rural health care workers in Hawaii (including, but not limited to, General Practitioners (GPs)) [13]. Dee and Blazek reported that, for rural physicians, communication with colleagues provided immediate, accurate and reliable answers and would invariably save the time and effort required to consult books and journals, even those readily available in the GP's personal library [14]. Work by D'Alessandro *et al.* on Digital Health Science Libraries for rural physicians may also have implications for veterinary information provision [15]. Finally, Thompson provides a useful review of the literature on the characteristics of information resources preferred by primary care physicians with reference to ISB/IU [16]. Outside of the medical domain, work by Ellis in academic and industrial environments provides a solid theoretical base for user study research, while Wilson has reviewed non-LIS research on IN and ISB [17, 18, 19, 20].

Methodology

Questionnaires (paper and email versions) followed-up by semi-structured interviews were the methods of data collection selected. The advantages of this approach include: a large geographic 'reach', comparably quick and cheap data collection and the opportunity to explore attitudes and opinions in more detail. The main disadvantage is the reliance on the goodwill of time-pressed and 'questionnaire fatigued' vets to participate.

Topics covered in the questionnaires and interviews were based on key issues identified in the literature review. A pilot sample of vets tested question design, survey layout and closed question responses. The RCVS Practice Database (RPD) was used as the sampling frame, from which a simple random sample was selected with the aid of random number tables. For the paper questionnaire, it was decided to use Valleley's sample size (n = 100) as a basis but with compensation for a longer survey (n = 149). The email questionnaire sample (n = 60) comprised vets whose email addresses had already been added to the RPD, or whose email addresses had been gleaned from three UK veterinary information websites [21, 22, 23]. The total sample population was thus 209. Pre-survey publicity, stamped addressed envelopes, official invitations, follow-up letters and a prize draw incentive (as recommended by Oppenheim [24]) were all used to boost the response rate.

The interview stage was conducted after the majority of questionnaires had been returned. Due to cost, time and field research constraints, seven interviews scheduled to last no more than an hour were undertaken. Candidates for interview were selected from respondents who had checked an interview consent box on the questionnaire on the basis of practice type, IT use, or geographical location.

Results

The final response rate was 39% (n = 82). This low rate is probably due to a combination of factors: survey length, subject matter and 'direct mail overload' affecting vets' attention and time. The email survey gained a slightly better response rate (42%) than the paper survey (38%). This may be because of the more 'immediate' nature of email, i.e. messages can be replied to quickly and easily with a

Table 1 Survey returns and interviews by practice type compared to UK totals

Practice type	Number (n)	Survey %	UK %*	Interviewed
Mixed	28	34	54	1
Large	5	6	1	1
Equine (100%)	3	4	3	1
Small	45	55	37	4
Poultry (100%)	1	1	0.3	0
Other	0	0	5	0
Total	**82**	**100**	**100**	**7**

*Source· RCVS Practice Database

click of a button (or, indeed, deleted!). Table 1 shows the final split of survey returns and interviews by practice type compared to estimated UK totals.

Computer usage

All respondents used a computer (rarely an Apple or Unix system). Laptop use was reported by just 17% of respondents and palmtop use 8%. The most popular computer activity was word-processing followed by accounts/invoicing and label printing (in that order). The least popular activity was website design or maintenance followed by veterinary note taking and database searching. If, for the purposes of analysis, email communication, internet searches, database searches, veterinary note taking and multimedia use are considered to be the principal examples of ISB from the list of activities given, the percentage of ISB activity out of total computer activity can be calculated. Thus ISB comprised just under a third (29%) of all respondent computer activity. There was no significant difference found between patterns of computer activity by practice type, size, or age group.

The Internet

'The biggest waste of time known to man', lamented one respondent. But the majority of vets disagreed – 66% used the Internet for work related purposes and a further 11% were using it for non-work related purposes. Of the remainder, there was a small group of vets who were 'Internet converts', i.e. they stated they would use it if they had the time. Only three respondents had no idea what use the Internet would be to their work. Despite the recent press attention on the comparatively high cost of using the Internet in the UK, compared to the USA, due to local call charges, such costs were not perceived as a barrier to Internet use with only one respondent mentioning them. The remaining 11% of respondents had other undeclared reasons for not using the Internet at present. Just over half the mixed animal practice respondents (54%) used the Internet for any purpose, as did 69% of the small animal practice respondents. All the equine practice respondents in the survey used the Internet and all bar one of the large animal practice respondents.

An analysis of Internet use by age reveals that it is the younger vet respondents who were least likely to be online (three out of seven), however this category is under represented in the survey. With this in mind, it was found that the 31–40 age group were the category using the Net the most (73%), followed by the 41–50 age group (67%) and the 51–59 age group (60%).

How did the sample of vets find out about Internet resources? Three main sources emerge (in rank order, most popular first): online sources (comprising surfing, browsing type activities, search engines, newsgroups and mailing lists), published literature sources (e.g. adverts, articles in journals or newspapers) and oral sources (colleagues and word of mouth).

And once these vets are online, what then? Some degree of caution should be exercised here as some of the categories of online resource provided were interpreted differently by different respondents. However, it is fair to claim that 83% of respondents who used the Internet for work went to professional institution websites at least occasionally, and 79% to veterinary information websites.

As for the categories of websites that were *not* visited, 78% of Internet respondents never used library catalogues on the Web. The word 'never' is perhaps too strong here as null responses are included: somebody ticking a 'never' box is obviously making a stronger statement than the person who leaves it blank, so 'do not use' is a more appropriate interpretation. But, semantics aside, the inescapable conclusion is that library catalogues are not being used, in spite of major recent capital investment in them by VIS.

However, mailing lists are used even less (85% of Internet respondents did not use them) and 74% of Internet respondents never used newsgroups, publishers' websites, or online newspapers in their veterinary work. The low reported use of online databases is also important as it shows that a veterinary equivalent is needed of the various MEDLINE gateways on the Net. Vets do use MEDLINE – this was the most cited example of an online database in the survey – but it is first and foremost intended for the medical profession and so its coverage is not specialised enough for veterinary users. Other frequently cited Internet resources were the Ministry of Agriculture, Farming and Fisheries (MAFF) site and the BSAVA web pages.

Although fully aware of the Net's potential as an information resource, it is fair to say that all the interviewees were at least frustrated by it in its current state, if not openly hostile. One saw an important role for the Net, thanks to its speed and immediacy, as the profession's preferred internal publishing/communication medium for position statements and professional issues. Speed of publication was also a key factor in another's appreciation of FEMS online as it facilitates current awareness. The Internet's immediacy was also valued by a third interviewee as it means there is the tantalising possibility of being able to satisfy any information need instantly.

Email

Around 79% of respondents had the use of an email address and the majority of them used email for work related communication. 64% of the survey's vets can be described as truly 'wired', using both Internet and email to help them with their practice work. An additional 16% used email and not the Internet in their work and 17% vice versa.

The majority of respondents seemed to be emailing other vets in the UK and 69% were corresponding with vets working in other countries. On the other hand, 91% of vets using email for their work did not exchange information with laboratories. Librarians did not fare well either with only 18% of the respective sample emailing them. As for clients and practice colleagues, it seems that just under a third of respondents were exchanging information with these categories of people. 62% of respondents who used email for work were communicating with between one and three types of recipient. 9% could be described as 'email addicts' as they were emailing seven or more types of recipient.

Leaving aside personal, non-work related information which predominated, clinical and case related information appeared to be the most popular type of information being exchanged, closely followed by practice management information. The relatively low amount of personal research information exchanged is noteworthy (7 responses out of 87).

Although, on the basis of the survey findings, it can be argued that email use is in its infancy in terms of ISB, some of the interviewees indicated that this situation might change rapidly. One stated that email use rises exponentially once it is used fairly regularly and becomes part of the daily routine. Another vet, a heavy email user, was of the opinion that email use saved paper and that having an email address was a vital part of client relations. Regular use gives rise to experimentation, explaining one of the fascinating interview findings – the use of flatbed scanners to exchange document images. One vet participated in a virtual vet community in which meeting proceedings and journal articles were exchanged via email and post. Another exchanged non-clinical information with his brother (also a vet) as well as maps and CVs with prospective visitors.

CD-ROMs

Of the 80 respondents who answered the question on CD-ROM usage in the past month, 12.5% indicated that they did not use CD-ROMs at all, while the majority (40%) had not used a CD-ROM in the previous month (19 of these were Internet users, 13 were not). The remaining 38 respondents were hardly heavy CD-ROM users as only 3 of them had used CD-ROMs 15 or more times in the previous month.

CD-ROMs were used mainly for 'staying current' (presumably as part of the vets' continuing professional development requirements) by 78% of regular users. This compares with 55% of regular users consulting CD-ROMs to answer patient care questions and 36% consulting them for personal research purposes. It was also found that regular CD-ROM users overwhelmingly (89%) preferred to search CD-ROMs by themselves rather than delegate their searches to somebody else.

Vetstream products dominate the veterinary CD-ROM market. Twenty respondents reported using CD-Canis, while twelve used the CD-Felis and five CD-Equus. Eight respondents used Vision. The response for MEDLINE (3) was probably not accurate, as many respondents may have considered it to be an Internet resource (e.g. via Grateful Med) rather than a CD-ROM or database.

Vetstream CD-ROMs were discussed in three interviews. Opinions were negative on the whole: the third interviewee found them to be expensive, clumsy to navigate through and conservative in that they do not go beyond the standard techniques that most vets know anyway. One vet obviously appreciated CD-Canis and CD-Felis as he worked essentially alone in a small animal practice and reasoned that the kind of information included on them obviated an online version. He did, however, acknowledge their limited scope/coverage. Another vet had looked at CD-Equus and found it to be expensive and, once again, not particularly detailed compared with journal articles. Another found the data compendia and formularies that are now available on CD-ROM to be more useful. Finally, one interviewee signalled an information need fulfilled particularly well by the CD-ROM medium – the communication of conference proceedings.

Are CD-ROMs threatened by the Internet? Although the third interviewee was pro-Internet, the second preferred CD-ROMs as she found that she wasted a lot of time using databases on the Internet. The fifth saw no need to tie up his telephone line when he had perfectly adequate information on CD-ROM. The sixth had suffered slow download times when online which made him appreciate CD-ROMs for their accessibility and portability as well as the functionality of certain interfaces. He was also concerned about general image quality. The fourth vet thought that CD-ROMs would replace paper journals, something that would facilitate indexing and searching. The first interviewee, however, was not impressed with the CD-ROMs she had seen due to their basic level. For her, books provided a viable alternative to the Internet.

The survey revealed four major reasons for CD-ROM non-use: lack of easy access, lack of time, lack of equipment and uncertainty over which databases to use. Reasons which were not considered important (and scored equally lowly) were: cost of searching, preferring others to search, dislike of computers, wrong kind of information and no access to journals once references found.

Preferred communication media

The closed question relating to this issue did not generate an entirely satisfactory set of data either as 16 of the 79 respondents gave multiple answers when only one response was required. However, the rank order of preferred media is unchanged after using either an elimination or compensation strategy in analysis: (favourite first) email, paper document in the post, Internet site, fax, telephone, floppy disk and audio cassette/other. Note that no distinction was made between different kinds of telephone. Fax and telephone are almost interchangeable in the rankings. For the 18 respondents who stated that neither they nor their practice had an email address, information in paper document form through the post was the preferred medium followed by fax and telephone.

Respondents' preferred means of obtaining information when working away from the practice was also explored in order to gauge the potential use of portable communication/information devices. Out of the 61 respondents who answered the question, 5 stated they would wait until they returned to the practice to get the information they needed. Fifteen used mobile telephones. An additional 31 vets said they used the telephone, which may or may not include mobile equipment. Interestingly, 7 vets stated they used books. Laptop use was reported by 8 respondents. Pagers were used by only 3 respondents.

Preferred information sources

However many respondents might be using the Internet and/or databases, these kinds of sources were not high up at all in the overall scheme of veterinary information seeking (Table 2). Journal articles, textbooks and conferences were the sources of choice for drug, diagnostic and therapeutic information. Use of other sources varied. It was found, for example, that promotional literature was considered to be a particularly important source for drug information (77% of respondents used it), as were company representatives (73%), and practice colleagues (68%). On the other hand, the relatively low overall use of databases and websites is probably a reflection of the quality of information provided and, especially in the case of websites, of a lack of credibility (i.e. sites that are biased and are neither peer-reviewed nor officially approved).

Information source use by practice type

A tentative analysis was made of variations in source use for each information type by small animal and mixed practice types. Mixed practice respondents used an average of three information sources for each information type, while the equivalent figure for small animal practices was six. Once again, these figures should be treated with caution because of sample size and response rate.

For drug information needs, it can be argued that respondents working in small animal practices were using practice records as an information source to a much greater extent than their peers in mixed animal practices (76% versus 32%). A similar relationship existed for 'other vets' as a resource (64% small animal use versus 39% mixed animal use). Small animal practice respondents were also using electronic information resources to a greater extent (+17% in the case of databases). On the other hand, mixed animal practitioners were making greater use of encyclopaedia and compendia than small animal practice practitioners and, while the

Table 2 Sources ranked by number of responses by information type

Information source*	Drug (n)	Rank	Diag (n)	Rank	Ther (n)	Rank	Total (n)	Rank
Journal articles	72	1	78	1	73	1	223	1
Textbooks	71	2	71	2	73	2	215	2
Conferences	63	3	71	2	69	3	203	3
Promotional literature	63	3	29	12	47	8	139	8
Company representatives	60	5	18	18	40	10	118	10
Practice colleagues	56	6	64	5	63	4	183	4
Training courses/workshops	48	7	61	6	57	6	166	6
Conference proceedings	47	8	51	7	50	7	148	7
Encyclopaedia/compendia	45	9	31	11	38	11	114	11
Other vets	45	9	65	4	61	5	171	5
Personal notes/files	43	11	43	8	42	9	128	9
Current awareness publications	31	12	35	10	33	12	99	12
Practice records	30	13	25	13	29	13	84	13
Other books	25	14	23	14	29	13	77	14
Databases	21	15	22	15	21	15	64	16
Internet sites	20	16	21	17	21	15	62	17
References/citations	18	17	22	15	18	17	58	18
Abstracting/indexing services	17	18	17	19	15	19	49	19
Laboratory	15	19	42	9	17	18	74	15
Annual reviews	10	20	10	20	11	20	31	20
Other	7	21	5	21	5	21	17	21

*Respondents could select more than one source overall finding that journal articles were the preferred resource for drug information still holds, textbooks, promotional literature and company representatives were of equal importance to mixed animal practitioners – more than the overall rankings in Table 2 would suggest.

There was a more defined 'information use hierarchy' for diagnostic information sources on the small animal side than on the mixed animal. In other words, mixed animal practice respondents were making greater use of a wider range of resources for this kind of information than their small animal practice peers. On reflection this is common sense – the wide variety of animal species seen by mixed animal practices, with all their respective conditions, must make calls on a wider variety of diagnostic information sources than is needed for a small animal practice. And it is this variety that may explain the greater use by mixed animal practice respondents of particular information resources such as training courses/workshops (+11%), practice colleagues (+8%) and laboratories (+13%) to help them cope.

Once again, small animal practices made greater use of databases for diagnostic information and abstracting/indexing services (33%:11% and 47%:36% respectively). However, there was a roughly similar use of Internet sites, albeit low in priority. One curiosity is that the small animal practice respondents reported significantly greater use of promotional literature (+23%) for this type of information. As suggested previously, promotional literature could be being used in conjunction with other information sources.

For therapeutic information, training courses/workshops featured more prominently in mixed animal practice respondents' routines (79% versus 67%), as did the use of practice records (46% versus 31%). On the other hand, current awareness publications, personal notes/files, references/citations, databases, abstracting/indexing services and laboratories were all used more (at least +10% for each one) by small animal practice respondents. The small animal practices in the survey used a wider variety of sources for therapeutic information than the mixed animal practices.

Information source use by practice size

A couple of observations can be made. Firstly, the apparently low use of electronic resources by 'super' (i.e. 11+ vets) practices, especially use of databases. This practice type in the survey did not use any abstracting/indexing services or annual reviews for any information type and reported very low use of references/citations. Further research is clearly needed here to confirm or deny these findings.

For one-vet practices, a clustering effect can be observed, especially for diagnostic and therapeutic information. That is, a small group of sources were used consistently more than others: journal articles, textbooks, conferences, other vets and conference proceedings. Only when drug information was needed were additional sources used to a similar extent: promotional literature, company representatives and personal notes/files. As for small (1-3 man) practices, it is fair to say that they generally used a wider variety of sources than other size practices.

Medium sized practices in the survey rated practice colleagues highly – the most useful information source for therapeutic information along with textbooks and second most useful source for diagnostic information. It is expected that this trend would also apply to super practices but their representation is too small in this survey to be certain.

Keeping up-to-date

Journals appear to be the most popular source for current awareness purposes (Table 3). However, it can be argued that meetings, courses and seminars all constitute a 'many to one' oral information source and consequently it is *this* kind of source that is the most used. Certainly, there is an overall balance between the use of published and oral sources. Responses such as 'CPD' and 'reading' conceivably included both journal reading, books and courses and so cloud interpretation. It is clear, though, that electronic information sources (and possibly one to one sources as well) were not regarded as

Table 3. Answers to open question – how do you keep up-to-date?

Classification					**Total**
Published	Journals			58	
	of which:	*Veterinary Record*	*10*		
		JSAP	*7*		
		In Practice	*5*		
		Other named examples	*15*		
	Books			12	
	Other literature			13	
	Reading (unspecified)			11	**94**
Oral (1 - ∞ - 1)	Meetings			35	
	Courses			20	
	Conferences			9	
	Seminars/lectures			6	
	Congresses			4	**74**
Oral (1 - 1)	Colleagues			11	
	Other vets			7	
	Other contacts			4	**22**
Electronic	CD-ROMs			7	
	Internet related			6	
	Audio cassettes			2	**15**
Miscellaneous	Other activities			4	
	CPD (unspecified)			7	**11**
	Total responses				**216**

being important, despite the fact they both offer the possibility of being the most current information sources of all.

Emergency sources

A similar conclusion was reached for respondents' optimum sources of information in emergency situations. The humble book (textbook or reference) was the main life support system (sic) with 45% of responses followed by practice colleagues with 18%. There is really little to choose between the remaining responses. Electronic information sources were clearly not regarded as being suitable for the job.

Barriers to IU

Time, cost and inadequate information skills training (in that order) were regarded as the main barriers to IU. Time barriers were twice as significant for respondents as cost barriers. This response was anticipated and an additional question gave respondents the opportunity to consider three possible solutions to the time problem, extracted from previous research. The results showed a very marginal preference for the 'circuit librarian' idea (+2) over 'onsite information intermediary' and 'personal intermediary at information service'.

However, the largest response was comprised of don't knows, no preferences and null responses (48%). This can be interpreted in several ways. Firstly, as a genuine expression of ignorance. Secondly, as a passive response, e.g. it is not for me to know the answers, others must present solutions to me. Thirdly, as an acknowledgement that there may not be a definite answer, with it being left to the individual vet's discretion to decide what action to take (if any) about time pressures. It should also be noted that respondents were given space to suggest their own ideas and only six chose to do so.

Problems with information sources

It is clear that currency of information (Table 4) was the most pressing problem with vets having to grapple with out-of-date information. One respondent pointed out that rapid changes in medicine and therapeutics, especially in disciplines such as anaesthesia, are rendering textbooks out-of-date as soon as they are published. Problems with availability and content here are also of concern. The former category encompasses occasions where colleagues or referral vets are unavailable at the time when the information need is most urgent, or where books have been left at home, or the information is at another practice in whatever form. Content problems refer to poor coverage of a topic, lack of detail on rarer topics, or conflicting/biased advice.

The fact that the cost of information does not appear to be a significant issue is worth retaining as is a particular example of a design problem – poor indexing. This refers not only to indexing of documents for a database or search engine but also for paper journals. Finally, at the risk of being provocative, it may be that the major problem was respondents' lack of information skills.

Evaluation of information

'Credibility of source' was by far the most significant evaluative criterion used with 40 out of 76 respondents allocating it first or second position out of the list provided. 'Personal experience' came second (n = 38) and 'discussion with practice colleagues' in third place (n = 25). 'Methodology used' and 'consultation with professional association' was hardly used as evaluating criteria.

Veterinary libraries

55% of respondents used a veterinary library and for 68% of these, RCVSWL was the most important library. No other veterinary library scored significantly. Of the 37 respondents who did not use veterinary libraries, 21 were Internet users, although that is not to suggest cause and effect. The opportunity was taken to ask these 37 non-users what (if anything) would prompt them to use a veterinary library

Table 4 ISB problems

Problem	**Example(s)**	**Number (n)***
Currency related	Out-of-date	19
Availability	Colleague busy	16
Content related	Biased, irrelevant	16
Physical	Bulky, hard to file	9
Design related	Poor indexing	9
Time consuming	-	8
Speed of access/delivery	Internet downloads	8
Finding the known reference	-	4
Lack of resource awareness	Where to look?	3
Too many sources	-	3
Other	Cost	3
Total responses		**98**

*Respondents could specify more than one problem and the majority of responses stated 'improve access via the Internet', i.e. accessing library resources online to download entire articles, reviews etc. This was a wish shared by the majority of library users

Veterinary libraries were primarily used for article/reference chasing (43% of the 69 responses to the question). The more general literature searching activity accounted for a further 19% of responses. Book or journal loans (26%) were thus not the principal library service used by respondents. One activity was mentioned that had not been previously considered – a minority of respondents were using veterinary libraries to evaluate textbooks before purchasing them.

A private affair?

One finding from the interview stage deserves further research – the strongly held view by an interviewee that any ISB should be done in private, that is to say, not in front of the client. This arose when discussing the potential use of portable data assistants when on call – the vet was adamant that consulting any source in front of a client was frowned upon as it was not seen to be 'professional'.

Conclusions

At first, it seems that the findings from this research merely confirm those published previously. Journals and textbooks continue to be the information sources of choice for practitioners in this survey's sample, just as they were some 20 years ago when Drake and Woods conducted their research. Despite advances in technology, books are still the emergency information source *par excellence* as (re)confirmed by Pelzer and Leysen in 1991. Even if more practices in this survey had computers (100%) than was the case for Valleley in 1993 (77%), the same conclusion applies now as did then – administrative tasks such as word-processing and accounting comprise the majority of computer activity, despite the arrival of the Net and email. What is more, the major barrier to IU identified by Raw – lack of time – is still as prominent as ever.

New ground has been covered, though, especially with regard to the Internet. More vets in this survey were using the Net for their work than were in Gerrard's survey (68% versus 28%) and there appeared to be less uncertainty about getting connected. Let it not be forgotten that more respondents reported using the Net for their work than reported using a veterinary library (by whatever means). This is in spite of the fact that RCVSWL and, recently, the Royal Veterinary College, have both invested in Sirsi's UNICORN library management system, which permits user friendly remote catalogue interrogation/interaction via the client World Wide Web browser.

Survey comments and interviews revealed that lack of publicity was partly to blame, but mainly it was the assumption that such systems fulfilled a proven practitioner need for bibliographic information via the Internet. In reality, it was found that there was a far stronger demand by VIS users and non-users alike for online full-text journal information and that an opportunity existed for the RCVSWL, as the 'profession's library', to assume a leading Internet role in filtering the useful from the useless on behalf of the profession. The principal of enhanced online veterinary information provision is thus the key service implication for RCVSWL that can be identified from current veterinary ISB/IU.

This should not be a surprise. It was anticipated in a 1994 article by the RCVSWL librarian at the time, Benita Horder [1, p. 556]:

> Future possibilities for our users may well include dial-up access, not only to our catalogue, but also CD-ROM databases. Electronic publishing is likely to increase over the next decade. Further specialisation and the need for continuing professional development will lead to increasing use. Individual access to databases is likely to become more widespread, but there will still be a need for the librarian as an intermediary and to provide the backing services of photocopying and lending.

Horder was writing just before the Internet became a public network and so was advocating direct dial-up as opposed to dial-up via a proxy, which is now commonplace.

However, her predictions then are now a reality thanks to the potential of the Internet. It remains to be seen if VIS (and more specifically, RCVSWL) are willing and able to tap it.

References

1. Horder, B. A profession's library. *Veterinary Record*, 134(21), 1994, 554–556.

2. Pelzer, N.L. and Leysen, J.M. Use of information resources by veterinary practitioners. *Bulletin of the Medical Library Association*, 79(1), 1991, 10–17.

3. Drake, M A. and Woods, L.A. An information service for practising veterinarians. *Bulletin of the Medical Library Association*, 66(4), 1978, 437–441.

4. Pelzer, N.L. and Leysen, J.M. Library use and information-seeking behaviour of veterinary medical students. *Bulletin of the Medical Library Association*, 76(4), 1988, 328–333.

5. Pelzer, N.L., Wiese, W.H. and Leysen, J.M. Library use and information-seeking behaviour of veterinary medical students revisited in the electronic environment. *Bulletin of the Medical Library Association*, 86(3), 1998, 346–353.

6. Bawden, D. and Valleley, C. Veterinary information: sources and use. *Aslib Proceedings*, 48(11/12), 1996, 266–270.

7. Valleley, C. *The provision of information for practising veterinary surgeons in Great Britain*. MSc in Information Science dissertation, City University, 1993.

8. Raw, M.E. Survey of libraries in veterinary practice. *Veterinary Record*, 121(6), 1987, 129–131.

9. Gerrard, B. Vets on the net (3): present and future. *In Practice*, November/December 1998, 602–606.

10. Biggs, H. *IT and the vet*. Produce Studies Research. [Cited in: News review, *Veterinary Times*, August 1999, 36.]

11. Gravois, S.L., Fisher, W., Patrick, S.C. and Bowen, D. Information-seeking practices of dental hygienists. *Bulletin of the Medical Library Association*, 83(4), 1995, 446–452.

12. Bowden, V.M., Kromer M E. and Tobia, R.C Assessment of physicians' information needs in five Texas counties *Bulletin of the Medical Library Association*, 82(2), 1994, 189–195

13. Lundeen, G.W., Tenopir, C. and Wermager, P. Information needs of rural health care practitioners in Hawaii. *Bulletin of the Medical Library Association*; 82(2), 1994, 197–205.

14 Dee, C. and Blazek, R. Information needs of the rural physician: a descriptive study. *Bulletin of the Medical Library Association*, 81(3), 1993, 259–264

15. D'Alessandro, D M., D'Alessandro, M.P., Galvin, J.R., Kash, J.B., Wakefield, D S and Erkonen, W.E. Barriers to rural physician use of a digital health sciences library. *Bulletin of the Medical Library Association*, 86(4), 1998, 583–593.

16. Thompson, M L Characteristics of information resources preferred by primary care physicians. *Bulletin of the Medical Library Association*, 85(2), 1997, 187–192.

17. Ellis, D., Cox, D. and Hall, K. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49(4), 1993, 487–501.

18. Ellis, D. Modelling the information-seeking patterns of academic researchers: a grounded theory approach. *Library Quarterly*, 63(4), 1993, 469–486.

19. Ellis, D. and Haugan, M Modelling the information-seeking patterns of engineers and research scientists in an industrial environment *Journal of Documentation*, 53(4), 1997, 384–403.

20. Wilson, T.D. Information behaviour an interdisciplinary perspective. *Information Processing and Management*, 33(4), 1997, 551–572.

21. Vet Web http://www.vetweb co.uk

22. Vet Net http://www vetnet.co.uk

23. VetIndex. http://www.vetindex.co.uk

24. Oppenheim, A.N. *Questionnaire design, interviewing and attitude measurement*. London: Pinter Publishers, 1992.

The following people are thanked for their invaluable assistance: Liz Armitage, David Bawden, Katie Wales, John Collins, Simon Jackson, Washiela Jacobs, Simon Jackson, Yioda Panayiotou, Brian Wales and Tom Roper.

The century of film: bibliographic control and legal deposit of the moving image

Stephen Pearson
British Film Institute National Library, Stephen Street, London W1P 2LN
stephen.pearson@bfi.org.uk

1895 saw the dawn of film – a new publication medium that we associate particularly with the century that has just ended. The survival of films and of information about them has been severely hampered in the UK by the absence of legal deposit. At the beginning of the new century, proposals to remedy that situation would not only benefit film but might, for the new digital media, help to prevent the loss of treasures on the scale of that suffered by film in the twentieth century.

Introduction

It seems appropriate at the beginning of a new century to look back at the state of our knowledge about the publications of the previous one. It seems particularly appropriate when, as in the case of film, the publication medium belongs almost entirely to the previous century – indeed it represented a revolutionary technological innovation at the start of the twentieth century. Such a review might help us to learn some lessons, and this is very appropriate at a time when methods of production and distribution of the moving image are undergoing dramatic changes, again because of the advent of new technologies.

Of the films that were produced in this country during the first thirty years of the existence of film technology, in other words the 'silent' era, probably no more than 20% survives. We have to use the word 'probably' because we do not have a complete record of all the films that were produced and we shall probably never be absolutely sure what has been lost. Even now occasional 'treasures' are rediscovered. In the absence of any formal requirement to register films, such as that established by the Library of Congress in the United States, the information that we do have about films produced during that period has had to be painstakingly pieced together from surviving copies of company film catalogues, from the surviving films and fragments of films, and from the trade magazines of the period, notably, for this country, *Bioscope and Kinematograph Weekly* and its various related titles. Of the films from the 'sound' era, usually regarded as having begun in 1927, perhaps 50% survives. For this period the records are rather better because, following the Cinematograph Films Act of 1927, it became a requirement to register the titles of films and their footage with the Board of Trade.

The development of film catalogues

The earliest attempt to provide a systematic record of films in this country dates back to 1934 when the first issue of the British Film Institute's *Monthly Film Bulletin* appeared. At that time the film records were very brief:

> **THIS IS THE LIFE.** British Lion. Censor's Certificate 'U'.
> Broad comedy with Binnie Hale and Gordon Harker.

The publication covered educational films as well as feature and short films. It also established the tradition of including not just British films but also those imported into this country. As the foreword to the first issue said:

> The British Film Institute, in deciding to publish a monthly bulletin of current educational and documentary films is endeavouring to meet a difficulty which has been experienced for sometime past by various organisations and individuals who have realised the possibilities of the film and who have been anxious to make use

> who have been anxious to make use of it in their work. They have been definitely handicapped by insufficient knowledge as to the extent and value of the material available on the market... [1].

The amount of information about each film gradually increased over the years until by the 1990s, when the *Monthly Film Bulletin* was incorporated into the bfi's other magazine, *Sight and Sound*, the records included full cast and credits, a plot summary and a review for films released in this country for cinema screening. *Sight and Sound* therefore continues an unbroken tradition going back to 1934 of cataloguing films, both British and non-British, that are released for the cinema in this country. To this needs to be added a whole range of film catalogues usually produced by enthusiasts, for example Denis Gifford's *Catalogue of British Films 1895–1985* [2], Rachel Low's *History of the British Film* [3], published in several volumes between 1948 and 1979, and David Quinlan's *British Sound Films: the studio years 1928–1959* [4].

However, by this time the 'educational' films had been dropped. The role of cataloguing films on a systematic basis, and more recently, videos and interactive discs, which are not intended for showing in cinemas, had passed to the *British National Film Catalogue* in 1963. This publication has had an almost unbroken run since then and is still being published as a collaborative venture between the bfi and the British Library. Now entitled *The British National Film and Video Guide*, it:

> lists films, videos and multimedia programmes with a substantial moving picture element offered within the UK for non-theatrical loan or purchase, i.e. copyright cleared for screening to non-fee paying groups in, for example, educational establishments, commercial or industrial concerns, and local authorities. Materials covered include educational and training films, independent productions, documentaries, television programmes and feature films. Cinema and TV commercials, amateur films and camcorder/home videos, in-house corporate communications and commercially-produced and released home videos are excluded [5].

As we have already seen from the foreword to the first edition of the *Monthly Film Bulletin* the bibliographic control of audiovisual resources in this country is closely allied to education and training. With the development of new teaching methods in the 1960s went a much keener interest in ways of involving and motivating learners. Audiovisual learning resources took on a much more important role because they could be used to provide an interesting and stimulating addition to the learning experience. These new approaches to teaching and learning permeated not only primary education, but also secondary, further and higher education.

In 1971 an organisation called the National Council for Educational Technology published a report entitled *Non-book materials: their bibliographic control* [6]. This report pointed out that:

> the cataloguing of educational materials in the various non-book media is becoming an increasingly important factor in the effective development and utilisation of these materials ... in many schools, colleges, universities, resource centres and other institutions on both sides of the Atlantic, the growth of collections of non-book media is now such as to raise serious problems of control and utilisation.

This, of course, was long before personal computer software, games and CD-ROMs had appeared on the scene.

The report went on to advocate the creation of a computerised database of non-book learning materials called the Learning Materials Information Store and proposed the use of the same computer system as was then used for the creation and publication of records for the *British National Bibliography*. The idea at this stage was to use the database to publish lists that could target the information needs of various groups of users (e.g. lists by subject area). A pilot project was set up to test the ideas presented in the report and this pilot led to the development of *HELPIS* (Higher Education Learning Resources

Programmes Information Service) – a printed catalogue of 300 audio-visual resources that used the same subject indexing system and a very similar format to *BNB*. It was restricted to materials produced by the higher education institutions themselves, and was intended to facilitate the exchange of such materials. It did not include commercially produced materials. It was published by the British Universities Film Council, which went on to become the British Universities Film and Video Council, an organisation that has been extremely active over the years in promoting the use of, and information about, film in higher education teaching and research. Several volumes of the *HELPIS Catalogue* were produced during the 1970s and early 1980s, including some special ones such as *HELPIS Medical*. Another publication, the *BUFC Catalogue of Audio-Visual Resources for Use in Higher Education* listed other materials not produced by the universities themselves. With the development of the British Library's BLAISE online service, the database became available online as a service called HELPIS-Line and is now published on CD-ROM as the BUFVC's AVANCE database. Amongst other interesting items, this CD-ROM also includes the North West Film Archive's *Catalogue of Black and White Silent Films 1896–1929*.

The *Monthly Film Bulletin* and *Sight and Sound* records and the records of the *British National Film and Video Catalogue* have been incorporated into the bfi's SIFT database, which was developed in the 1980s and now contains over 600,000 film and television titles not just from this country but from all over the world. The database software does not at present enable it to be made available via the Internet but, as a step towards providing access from beyond the bfi National Library's Reading Room in central London, it is being made available via telephone dial-up from a number of centres across the country – the first terminal has recently been installed in the National Library of Wales in Aberystwyth. The catalogue records for the holdings of the National Film and Television Archive are also in SIFT. This includes films from before 1934 that are in the Archive. In addition, considerable retrospective work was done on the British film records as part of the bfi's contribution to a European filmography [7], a project that was, sadly, aborted.

Bibliographic control and legal deposit

A discussion of all these sources raises the inevitable questions what is it all for?; why do we need bibliographic control? Of course the obvious answer is so that we have a comprehensive and accurate record for posterity of our nation's film heritage. Ideally, the existence of a record for a film would also provide an assurance that the film itself existed in a national archive, in the same way that a record in *BNB* represents an assurance that a book is available in the British Library. This brings us inexorably to the question of legal deposit. In 1986 Line spoke about the principle of UAP – the universal availability of publications – whereby 'each country should be able to supply its own publications to whoever wants them'. Line's view was that legal deposit is essential to achieving this objective.

> 'An extension of legal deposit to all British publications is something that should be strongly urged. It is totally illogical that only that part of Britain's output of recorded knowledge should be deposited that happens to be in printed form...' [8]

Whether or not it is necessary to have legal deposit in order to establish bibliographic control is a question for debate, but in the last analysis the answer probably lies in the success of the relationship between legal deposit of books and the fact that we have in the British National Bibliography such an excellent record of the nation's book publications. Another illustration of the value of legal deposit is the discovery in recent years of the film copyright collection at the Public Record Office, Kew [9]. Under the terms of the Fine Arts Copyright Act of 1862 it was possible to register copyright in a photographic image at Stationers Hall in London. In most cases the applicants chose to include a copy of the image as part of the registration process. This procedure was followed until the passing of the International Copyright Act in 1911, and amongst the tens of thousands of photographs were recently discovered a few hundred records for motion pictures. These included single frames or small strips, which, in most cases, are the only remaining visual record of the films. The collection has revealed the existence of previously unknown early British filmmakers. Exciting as this discovery is, it needs to be compared with the many thousands of complete paper prints of films that have survived in the Library of Congress in the USA, deposited there in order to register copyright.

Some problems of film cataloguing

In addition to the assurance of survival that legal deposit provides, however, there is another very important link between legal deposit and bibliographic control. As far as film is concerned there is no doubt that, in the vast majority of cases, the credit and other information on the film itself should, ideally, be the source of the filmographic record, in the same way that the title page of a book is regarded as the authoritative source for the bibliographic record. However, the situation for film is more complex. The bibliographic record for a book relates to all identical copies of that edition of the book. A different edition will normally have its own separate bibliographic record, which clearly identifies it as different in terms of its pagination, illustrative content, date of publication etc. There is a parallel with films except that the 'editions' are not so clearly identifiable. Films can have changes made to them for all sorts of reasons. They may have sections cut from them as a result of censorship. The censorship cuts that are made for cinema release may differ from those made for the video release, or for showing on television. Films may have different language soundtracks dubbed onto them for release in different countries, or they may be subtitled. Sometimes they can be re-released with extra footage added to them, as happened in the case of the 1989 re-release of Lawrence of Arabia. In many cases they may simply have had sections cut from them as a result of the damage caused by repeated passage through projectors. Films can be released in many different formats – 70mm, 35mm, 16mm film, VHS videotape, videodisc, DVD etc. etc. There is a parallel here with the hardback and paperback formats in which books appear, except, of course, that there is a far greater variety of movie formats, and the problems are compounded by the fact that the different formats are often combined with changes of content. By the time a much-projected film arrives in a film archive it can prove very difficult to establish exactly what resemblance it bears to the 'original', especially given the fact that the original may itself be very ill-defined. None of this information is normally shown in a catalogue record which records information about the work rather than about specific copies or versions, although the bfi's SIFT database does record different running times for films and changes in content where this is known [10]. At the very least, the deposit of a copy of the film as first released could lead to the establishment of a 'benchmark' catalogue record against which subsequent variants could be compared.

There is also the problem of establishing the nationality, especially of feature films. There are internationally agreed rules for film cataloguing compiled by the Fédération Internationale des Archives du Film. These state that the country of origin of a film is 'the country of the principal offices of the production company or individual by whom the moving image work was made' [11]. However, there is much controversy and debate about this. Some would argue that the nationality of the majority of the cast and crew of a film and the shooting locations should play at least some part in determining nationality. Others argue that it should be decided by the main source of finance – something that can be difficult to establish in these days of international co-productions in which companies from several different countries provide the investment. Attempts have even been made to devise a system to allocate points to a film according to various combinations of these factors. This would involve a considerable research task, especially the verification of such information as the nationalities of participants. Politics and national pride further complicate the problem. The recent excitement in the popular press about British winners at the Oscar ceremony serves to disguise the fact that, in reality, British film-making and distribution is still dominated by Hollywood – the fact that *The Full Monty* was directed by a British director, shot in this country, and had a largely British cast and crew, should not mask the harsh economic reality that most of the profits go to the American company that made it. For this reason it is recorded on the bfi's SIFT database as a US/GB co-production, rather than as GB or GB/US. However, we digress. The main point here is that the information on the film itself, and especially the copyright statement, is an invaluable indicator to the film cataloguer about the country to which a film 'belongs', and is particularly useful in helping to sort out the complexities of the countries to which co-productions 'belong'.

Copyright and film archives

In countries such as the UK where there is no legal deposit the relationship between archives and film producers has often been developed through painstaking negotiation and the development of mutual trust [12]. Film producers were often worried about the potential loss of control over their material that would follow on the deposit of their films. Clearly one of the main purposes of an archive is to provide access to films that it would not otherwise be possible to see. Individual viewing of a film is not normally considered to be an infringement of copyright – it does not constitute making the material public. However, as soon as an archive shows a film to more than one person at a time it enters a new realm as far as copyright is concerned. This is because 'screening for a public audience, even a select one, is considered exploitation, whether or not we charge admission fees, and such exploitation requires the permission of the copyright owner' [13].

Another area of copyright that concerns film archives is the question of how copyright law relates to the copying of films for preservation purposes [14]. In most countries copying for preservation purposes without the permission of the copyright owner would be a breach of copyright law. The problems become even more complex when restoration takes place, because this may involve taking footage from several different sources with the result that the restored copy may be very different from any of the extant preserved copies.

Moves to establish legal deposit

For these reasons archives in most countries have had to tread a very careful path in order to fulfil their obligations to collect, preserve, restore and provide access to their collections. There have been several attempts in the UK to extend legal deposit to audio-visual materials. In 1969 Dr David Kerr introduced a Private Member's Bill in the House of Commons, which received all-party support but which failed to receive financial support from the then Labour government. Subsequently the Wilson Report of 1976 and the Interim Action Committee on the Film Industry in 1978 both failed in achieving legal deposit. As Clyde Jeavons, former Curator of the National Film and Television Archive, said in 1993 'The Archive has continued to rely almost solely on the skilfully nurtured goodwill of the film, television and video industries for the voluntary deposit of moving image material in the NFTVA.' [15] Jeavons was putting forward the case yet again for legislation which would require producers of films made in the UK, and importers or distributors of non-British films for showing in the UK, to deposit on request an unused original format print. This legislation should also establish the bfi's rights to 'preserve, copy and provide unhindered access to the materials for bona fide educational purposes'. In addition to clarifying the legal position with regard to the activities of the National Film and Television Archive, it was also argued that it would make much better economic sense for the Archive to receive an undamaged print, than for it to continue to have to carry out costly restoration work on whatever used and damaged material it is able to acquire. It was also argued that 'film' should be extended to include a wide range of audio-visual material, including newer technology such as computer games.

Four years later, in 1997, a government consultation paper *Legal deposit of publications* [16] contained a chapter on 'Extending legal deposit to new publication media' in which it was stated that 'The Government's view is that some means must be found of ensuring the continuing comprehensiveness of the national published archive, and that legislation is likely to be required to extend the legal deposit system to publication in new forms.' These 'new forms' included film, CD-ROMs, sound recordings, online publications and microform publications.

In 1998, the new Labour Government set up a Working Party on Legal Deposit, the remit of which was:

> to advise on how an effective national archive of non-print material might be achieved .. to draw up and agree a voluntary code of practice to achieve deposit of electronic and microform publications until such time as the Government may decide to introduce legislation, and to ensure that such arrangements are compatible, where appropriate, with the existing arrangements for the voluntary deposit of films and sound recordings [17].

Its conclusion about the first of these questions was that 'only a system of legal deposit will be adequate to secure a comprehensive national published archive'. In view of the fact that it has taken over a century to arrive at a situation where there is at least the prospect of legal deposit for film, it is encouraging to note that the report recognises the need for a legislative framework that would 'empower the Secretary of State, after appropriate consultation, to declare, from time to time, publications in specific new media to be subject to ... the obligation of legal deposit'. In the case of British films, the report proposed that 'a mint print of the authorised premiere version' should be delivered within thirty days of first showing and that 'the original negatives or acceptable alternative pre-print materials' should be deposited after three years. For 'other films shown or distributed in the UK' a 'best copy of the film in the form first shown in the UK' should be deposited. Amongst the benefits of this system it was noted that materials would be preserved in optimum storage conditions, at no cost to the producer, that expensive restoration would become unnecessary, that producers would have access in order to make new copies and that the work would survive 'for future generations to appreciate and enjoy'.

The report is to be commended for its unequivocal espousal of the principle of legal deposit, but in some respects it represents a stopgap because the Government could not find time in its legislative programme to implement the recommendations of the 1997 consultation paper. As the Secretary of State, Chris Smith, wrote 'as a result I have had to look at alternative means of plugging the rapidly increasing gap that is developing in the national collections in respect of non-print material'. These 'alternative means' are the continuation of existing voluntary arrangements for film and sound recordings and their extension to electronic and microform publications.

Conclusion

With over a hundred years of production of film and recorded sound behind us, and the dawn of a new age of digital media, we are probably closer than we have ever been to the establishment of a legal deposit system that could ensure that the preservation of our audio-visual culture is no longer left to chance. The development of online catalogues of these resources, available on the Internet and based on information taken from the films themselves, opens up the parallel possibility that no-one need be deprived of accurate information about them.

References

1 Foreword *Monthly Film Bulletin*, 1(1), February 1934.

2. Gifford, D. *The British Film Catalogue 1895–1985· a reference guide*. 2nd edition. Newton Abbot: David & Charles, 1986.

3. Low, R. *History of the British Film 1896–1939*. 6 vols. London: Allen & Unwin, 1948–1979.

4 Quinlan, D. *British Sound Films: the studio years 1928–1959*. London: B.T Batsford, 1984

5. Preface. *British National Film and Video Guide*. Boston Spa: British Library, 1999.

6. Gilbert, L.A. and Wright, J.W. *Non-book materials: their bibliographic control: a proposed computer system for the cataloguing of audio-visual material in the United Kingdom*. London: National Council for Educational Technology, 1971.

7. Nowell-Smith, G. A European filmography. *Journal of Film Preservation*, 47, 1993, 4–11.

8. Line, M. Access to the nation's resources of non-book materials. *Audio-Visual Librarian*, 12(2), 1986, 63–68.

9. Brown, R. The British film copyright collection. *Journal of Film Preservation*, 54, 1997, 50–51.

10. For a fuller discussion of these problems see: Nowell-Smith, G. Filmography. *Screen*, 32(4), winter 1991, 452–455.

11. Harrison, H.W., ed. *The FIAF cataloguing rules for film archives*. Munchen, London, New York, Paris: K.G. Saur, 1991.

12. Houston, P. *Keepers of the frame: the film archives*. British Film Institute, 1994. [A fascinating study of the development of the film archive movement.]

13. Blotkamp, H. The symposium on archival rights in Jerusalem. *Journal of Film Preservation*, 52, 1996, 28–35.

14. For a detailed discussion of the copyright issues see: Henry, M. Copyright, neighbouring rights and film archives. *Journal of Film Preservation*, 49, 1994. 2–9.

15. Jeavons, C. The permanent way. *Journal of Film Preservation*, 47, 1993, 16–18.

16. Great Britain. Department of National Heritage. *Legal deposit of publications: a consultation paper*. London: Department of National Heritage, 1997.

17. Great Britain. Department of Culture, Media and Sport. *Report of the working party on legal deposit*. London: DCMS, 1999.

P 35,540

The European information needs of secondary school teachers in Scotland: recent developments in the provision of information to schools and colleges

Gillian Conroy
Belfast Public Libraries, Royal Avenue, Belfast, Northern Ireland, BT1 1EA

Susan Parker and Sylvie Davies
School of Information and Media, The Robert Gordon University, Garthdee Road, Aberdeen, AB10 7QE
s.parker@rgu.ac.uk s.davies@rgu.ac.uk

European Union (EU) education policy encourages the integration of a European dimension across the curriculum to prepare young people for participation in the EU. This article discusses the findings of a survey investigating the European information needs of secondary school teachers in relation to the teaching of the European dimension. Interviews were undertaken with a small sample of teachers and school librarians in selected schools of Aberdeen City and Aberdeenshire. The article discusses three main issues: inconsistency in the implementation of the European dimension at school level, ambivalent perception of what it actually means, and need for appropriate resources to enable teachers to include the European dimension in their teaching. The European Commission Representation Office in the UK has responded to the needs of the school sector by establishing a network of European Resources Centres for Schools and Colleges. Further research is required to assess the impact of these developments.

Background: EU education policy and the provision of European information to schools

European Union education policy within the principle of subsidiarity

The Maastricht Treaty provides a legal framework for the co-ordination of a Community education policy, based on a Resolution in 1988 [1], to develop a European dimension at all levels of education. This would seek to strengthen a sense of a European identity in young people and would equip them to take part in the economic and social development of the European Union. Accordingly, in 1991 the European Ministers of Education agreed 'that the basic values of political, social and individual life which underline the educational process must be viewed in the framework of a wider European community of peoples and states' [2].

According to the principle of subsidiarity whereby action is taken at the level where it can be most effective, the ultimate control of the policy is passed to individual national and local governments of each member state. The British government responded to the call of the Resolution in a document published in 1991, recommending a legislative framework to support the implementation of a European dimension while ensuring that ultimate responsibility should lie with the local education authorities [3]. More recently, and as one of the outcomes of the UK's Presidency of the EU, in early 1998 the government reviewed the issues of the European dimension in education in relation to its implementation, its content and its evaluation [4, p. 1].

The response from the Scottish Office has been to develop its own strategy within the framework of its policy document: Scottish education and the European Union – policy, strategy and practice [5], which is also supplemented by guidelines issued by the Scottish Council on the Curriculum (SCCC) [6].

Aberdeen City and Aberdeenshire Education Departments have issued a policy statement in relation to the European/international dimension in education [7]. Their aims are based on Article 126 of the Maastricht Treaty and the Scottish Office Report, and include recommendations to make young people aware of opportunities and challenges at the European level and to encourage understanding of the diversity between the member states. These aims can be achieved by:

- promoting international links;
- integrating a European dimension into the curriculum;
- providing appropriate curricular guidelines, information and resources, etc.

European information provision for schools in Scotland

As part of the European Commission's Information and Communication policy, a network of European Information Relays has been established in each of the member states to provide access to European information at the local level. In the UK, the Representation Office of the European Commission has been proactive in its approach to a decentralised information policy: its website [8] lists local sources of EU information on a regional basis. The network of information relays aims to provide European information to serve the needs of specific sectors, such as European Documentation Centres who serve the higher education community and Euro Info Centres who respond to the information needs of the business community. In Scotland, the Commission supports the members of the Relay through a recent development: the Scottish European Relays Network (SERN) [9].

However, from the mid-1990s, the European information needs of the school sector was beginning to be acknowledged:

> With 70% of written and 35% of telephone enquiries received by the London office of the European Commission coming from students, pupils, teachers, the current priority is to set up a system by which EU information can be more readily accessible at the local level [10, p. 30].

Subsequent meetings were held among Commission officials, the Department for Education and Employment, and representatives from local education authorities with the aim of tackling the information needs of education [11]. At the Second Annual Conference of UK Relays in 1996, the establishment of an Education Relay was formally announced [12]. In the following year, and on Europe Day, 9 May 1997, the London Office of the European Commission was the venue for the launch of European Resource Centres to meet the information needs of schools and colleges [13]. However, these Centres were initially set up to cover England and Wales only.

These developments had not reached Scotland as it was felt that services already in place could meet the needs of the school sector. However, it is clear that there still remained a gap within Scotland in terms of a service that could specifically target and meet the needs of this sector. In 1998, teachers in Aberdeen City and Aberdeenshire relied for European information on an eclectic range of information providers, which are presented below in Table 1.

It is clear from Table 1 that there is a gap in terms of an information service that is designed for, and would respond to, the European information needs of schools and colleges. This is reinforced in the literature by Morton, who states that:

> ...whilst these [information relays] do an excellent job in serving the needs of their own sectors – the general public and higher education etc, they are not ideally suited to the needs of schools and colleges [14, p. 14–15].

Table 1: Organisations/information services consulted by secondary school teachers

Organisations/services	**Role/ information**
National level	
The Eurydice Unit [15]	The role of Eurydice, the Information Network on Education in Europe, is to promote the exchange and the production of reliable and comparable information on education systems and national polices in the field of education in Europe.
The Central Bureau for Educational Visits and Exchanges [16]	The British Council promotes educational and intercultural linkages between schools around the world through the Central Bureau for Educational Visits and Exchanges.
Scotland	
The Representation Office of the European Commission in Scotland [17]	Supports a network of European information relays in Scotland: Scottish European Relays Network (SERN)
Eurodesk [18]	Eurodesk is a European Network for the dissemination of European information and for the provision of telephone enquiry answering services, at national or regional level, for young people and those who work with them
Local level	
European Documentation Centres (EDCs) [19] and the European Public Information Centres (EPIC) [20]	EDCs meet the information needs of the higher education community, whereas EPIC's provides European information for the general public, via public libraries.
Local Authority	Both Aberdeen City and Aberdeenshire have international departments that support participation in EU education programmes. Education Officers provide practical and factual information, such as finding partners for projects.

In Scotland, Eurodesk is recognised as being the key service for providing information to young people and, in particular, providing information on funding that is available to the education, training and youth sectors. Eurodesk is funded by the Directorate-General for Education and Culture and its services include:

- a free telephone enquiry answering service, financed by subscription fee from enquirers;
- regular information and newsletters on European issues and funding programmes, as well as an email alerting service sending information to relevant groupings to apply for funding;
- a database available on the Internet.

In relation to the European dimension in the curriculum, Eurodesk do very little to encourage its implementation, and this is due to the fact that the focus of their service is to provide information and assistance on funding programmes available in the EU. However, in terms of recent developments in this area, the network of European Resource Centres for Schools and Colleges were extended in 1999 to cover both Scotland and Northern Ireland. As a result, Eurodesk, based in Edinburgh, now share the same office with the newly established European Resource Centre for Schools and Colleges. Prior to this development, and due to the fact that a gap existed in Scotland for this sector, a survey was carried out to investigate the European information needs of secondary school teachers in Aberdeen City and Aberdeenshire.

The survey: methods and results

The introduction of a European dimension in education requires the production and the provision of information whether it relates to the promotion of international links, or to the development of a new curriculum. A literature search has identified only one article discussing the use of EC materials in schools and colleges [14]. The article discusses a four-year research project with the view to producing 'an educational product or products supporting post-16 courses initially in the Merseyside area, but with potential for development on a broader scale'.

Given the absence of systematic research in Scotland in this area, a study was carried out during spring 1998 among six secondary schools in both Aberdeen City and Aberdeenshire. The present article discusses the findings, which focused mainly on the information needs of secondary school teachers in their effort to include a European dimension in their teaching activities.

The research method on which this survey was based involved two sets of semi-structured interviews. The first set focused on 15 teachers of a broad range of subjects (science, social science and English) to elicit the nature of the European dimension taught in their particular subject, the types of resources used and how the teachers accessed the material. The second set of interviews were undertaken with 5 school librarians and sought to identify the types of Europe related resources that are held in the library. In total, 20 interviews were carried out.

The European dimension in secondary schools

> All teachers have a part to play in shaping pupils' awareness and readiness to participate in this broader setting by integrating the European dimension within the curriculum where possible and also providing progressive opportunities to experience the reality of Europe through communication and contact [5, p. 44].

The International/European Education Officers of Aberdeen City and Aberdeenshire Councils support EU education policy by setting up school partnerships with other European schools as part of the Socrates-Comenius[1] initiative. However, the results have shown that at school level, the success of its implementation relies on the interests and enthusiasm of individual head teachers of schools and teachers.

Two schools were in the process of setting up new structures to formalise the European initiatives already operating in the school. One school had already developed successful links with other schools in Finland, Norway, Sweden and Germany, and was in the process of setting up a European Committee to develop these links within a more definite framework of action. Another school had established a European Awareness Committee and also planned to run a cross-curricular European Awareness Week. Both of these developments originated in the modern studies and modern language departments with the hope that it might include other departments, such as home economics, art and music. Where these types of activities exist, European issues became an important part of a particular subject area or of a school's activities. However, it must be stressed that none of the six schools who co-operated in the study had a written policy towards integrating any aspect of the European dimension.

The results from the interviews with teachers indicate that inclusion of the

---

[1] 'Socrates [21] is an EU action programme for transnational co-operation in the area of education. The Comenius chapter of Socrates aims to provide support for a number of initiatives designed to enhance the quality of school education and promote the development of a stronger European dimension in the educational process.'

Table 2. Subject areas in which the European dimension is included in the teaching

	European Dimension	**No European Dimension**
Scientific Subjects (biology, physics, general science)	✓	✓✓✓
Social Science (modern studies)	✓✓✓✓✓✓	
English	✓	✓✓✓✓

European dimension depends upon the initiative of class teachers and that it can take many forms. The European dimension was fully integrated in the modern studies curriculum only as illustrated in Table 2. In English and science subjects, teachers were generally not aware of a European dimension and how it could be incorporated into their subject.

From the results, only one biology teacher stated that s/he was involved in a Comenius Action 1 Programme[2]. This involved an environmental monitoring project which had been integrated into standard grade biology where data was exchanged with schools in Norway, Denmark, Ireland and France. The links have been developed to such a point that pupils would also communicate with one another on a social level. Overall, the pupils responded in an enthusiastic manner to the project and the teacher had encountered no difficulties with the incorporation of this project into the subject.

One English teacher indicated that the European dimension was included as part of the subject. The examples provided included looking at the position of other European countries, by investigating the EU in discursive essays on a wide range of issues, such as gun laws or euthanasia, and in the literature through French short stones in translation. It is worth noting that this teacher viewed these activities as a European dimension, where others did not. Clearly, there is confusion about what the 'European dimension' means in concrete terms.

In the subject of modern studies, the standard grade syllabus includes a module on 'International Relations', looking at conflict and power in relation to the European Union, NATO and the United Nations. Other modules focus on a comparison of lifestyles in different countries of the EU, looking also at the educational and employment opportunities available and the impact of the EU on the local area. Within the syllabus there are options to look at the structure of the EU and the various institutions, but one teacher pointed out that it was too complex for many pupils who have difficulties in understanding how the EU operates. For this reason teachers tend to focus on issues that relate to the social and cultural aspects of the EU.

Among the seven teachers who did not include a European dimension in their subject, four taught English and three taught scientific subjects. The main reason for not including a European dimension was the lack of time, how it would fit into the course, and lastly, how to locate relevant material to resource a European element. Furthermore, it was something that teachers had never considered as being part of the English or science remit. One teacher summed up the general opinion:

> But it certainly isn't anything that we're aware of, that we're making any particular moves towards ... it isn't something, to be honest, that comes to mind as being one of our priorities...

The results indicate that these teachers were obviously unaware of how the European dimension affected them. In only one case was the European dimension included in the teacher-training course which covered one or two examples of education systems in other European countries. The results also indicate that they were not informed of European issues that relate to their particular subject area. This was reinforced by one teacher who stated that s/he has not been sufficiently informed about this whole area:

> There's also an awareness of one's own personal ignorance. It's opening

---

[2] Action 1 is the promotion of school partnerships and projects.

> up a whole new area where you may have sketchy knowledge, but may not know enough to want to put yourself on the line.

In other cases, the European aspect has emerged within discursive essays when looking at the European perspective on a particular issue. Teachers described this as 'incidental', but as it does promote knowledge and an understanding of a wider Europe, it can be described as a European dimension. Here again, it seems to indicate a lack of awareness of what the European dimension actually entails.

European information needs of secondary school teachers

As demonstrated above, the notion of a 'European dimension' is multifaceted and the interpretations on how it should be integrated into the curriculum vary greatly, hence the diversity of sources used. From the analysis of the interviews, this section will highlight the specific European information needs of teachers in relation to the currency, content, level and format of the sources consulted.

Currency

Current information was felt to be vital when including a European dimension, especially in the teaching of modern studies, which focuses on current affairs.

> Things change so much in modern studies and we have to accept that. What we can do is get material the core part of which stays the same and then we can add onto the bits that change...

For this reason, the Internet and the media are the most useful sources of information as it was felt that videos, CD-ROMs and textbooks date very quickly and are often very expensive to up-date. Eurodesk and the European Commission were judged as 'very good' in terms of providing up-dates on European issues.

Content and level

Issues of material content constituted the overriding theme of teachers' responses. Most teachers stated that they would welcome more information on the social and cultural aspects of the EU, as pupils find learning about the institutions and the operational aspects of the EU too difficult and related to that, the fact that they were felt to be 'not at all interesting'. Therefore teachers would prefer to receive information on the lives of young people in the rest of Europe. The teachers stated that they would like data about school-age pupils so that comparisons could be made with the UK and other member states.

One teacher suggested that access to junk mail or market research would be useful. Another felt that a video, which showed the lives of children in each of the member states could be used by several departments within the school such as home economics, art and music.

A recurring problem is the need for resources that apply to all ages, as it was felt that there seemed to be a lack of suitable information for standard grade pupils. The level of publications was recognised as being a key issue, as it was felt that there was a need for resources to be tailored to both pupils and teachers. At present, teachers often have to re-write material and compile booklets and worksheets using activities and pictures from other sources, so that they can guide pupils through the subject. It was also felt that support from the teacher in the form of directing pupils to specific pages of books or parts of a CD-ROM, is essential.

Sources and format

The Internet is utilised by all teachers due to the fact that it provides up-to-date material and is popular with the pupils. However, not all schools have the facilities to allow widespread access. The Internet is the principal source of information for the standard grade biology Comenius project in one school, as the project has its own website based in Norway through which project data can be exchanged.

All of the teachers refer to CD-ROMs as a common format and the most widely used title is 'Europe in the round'. This is a commercially published encyclopaedic CD-ROM, which provides factual information on

European countries and the European Union, as well as information on job and educational opportunities. Teachers will also use video material, such as the video produced by a local MEP on his role in the European Parliament. However, this source targets primary seven and first year secondary school level, and therefore has limited use. Other titles available in most of the schools include 'Dial 010 for Europe' and '1992' which contains information associated with the Single European Act. Teachers find this format particularly useful as it tends to hold pupils' attention better than textbooks. Although popular, currency of the information contained within these sources is an issue.

The press and television is another widely used format for current information on European affairs. It was felt that the local press is particularly useful as it enables pupils to read about the effects of the EU on their local community. National newspapers also provide up-to-date information on subjects covered in class, but these can be sometimes too complex for the younger pupils. Television programmes are also used to illustrate European issues:

> ...what I would do is just show them – not an hour's Panorama, but 10 minutes that was relevant to us. Over a period of time I would maybe show all of it.

In this way, pupils receive current information and several interpretations of an event.

Five out of the six modern studies teachers use the ' ..huge amounts of information available from the European Commission' for general policies and issues. The most useful publications include *Europe on the move, Portrait of Europe, Exploring Europe, Europe today and What exactly is Europe*?

These booklets are designed specifically for the school sector. They include colourful graphics and the contents are usually more accessible for school pupils. However, the majority of these titles have been designed for primary schools. Teachers accept that these constitute a major improvement in the resources that have been published by the European Commission in the past:

> These new ones are much more colourful and presented in a much better fashion and hit a much better age-level.

One of the teachers who was not currently incorporating a European dimension in her subject, suggested that a newsletter format covering all subject areas would be useful and cheap to produce.

General commercial textbooks are also consulted but it is not a source identified as particularly useful for teaching the European dimension. It was also felt that other than the textbooks written specifically for the modern studies syllabus, general textbooks on the EU are too complex, aimed at too high a level or that they 'try to be all things to all people'. In other cases, these texts do not cover the EU in enough detail or do not meet the specific information needs that teachers have.

Access to European information: internal and external

Internal access: the library

Interviews with the school librarians indicated that only a minimal amount of European information was stocked in the school library because it was not often requested. However, one respondent had noticed an increase in demand for European information over the past two years. The latter also suggests that the inclusion of the European dimension is linked to the interests of individuals or departments within schools.

When European resources were used in the library they tended to be linked to requests for information on individual countries in the format of textbooks, encyclopaedias or CD-ROMs, which are suitable to all subjects and to all years. These sources contain factual information on a given country and are regularly used by two departments, including modern studies and geography.

A *'European Union information binder'* was recently sent to all local schools by the Schools Library Service, which contains up-to-date pamphlets that cover all EU related subjects. The binder is also regularly up-dated by the Schools Library Service.

Two school librarians had developed

additional folders on EU information with newspaper cuttings and articles. A further school was currently building up 'Topic Boxes' in the library, which contained articles selected by the school librarian and teachers, and these would be used in all subject areas. None of these 'topics' specifically addressed Europe, but it was expected that the EU would be referred to within the subject areas covered. Examples of topics include the environment, the elderly and agriculture. Within the school library there were also textbooks on the EU, its history and policies. European issues would be covered as part of other subject areas, such as agriculture and the economy. However, the overall opinion of the school librarians was that 'there is not a huge demand for European information'.

External access

According to all teachers interviewed, including those with no recognised European dimension in their subject, the media was cited as the most popular source for unofficial information. Journals and professional association newsletters also contain some European information, especially for the science subjects, but these were not used in the classroom due to the fact that the language is felt to be too difficult for the pupils.

The European Commission in Edinburgh was the most popular source of official information for the majority of modern studies teachers. Other external sources mentioned by the teachers, included the Central Bureau for Educational Visits and Exchanges, NATO headquarters and the Scottish Office. Only two modern studies teachers used Eurodesk whilst the other teachers stated that they were unaware of the services they provide.

In Aberdeen, teachers would rarely, if ever, use the European Documentation Centre and the European Public Information Centre and few were aware of their services. Those who have used these sources said that the information available was too detailed for school pupils. This is to be expected as EDCs and EPICs were not established to meet the needs of the school sector.

Teachers were also asked if it would be useful to have access to a European information provider for the school sector in Aberdeen. While some felt that it was simple enough to telephone or email the European Commission or Eurodesk in Edinburgh, others agreed that a local service would be preferable. One teacher also stated that a satellite version of Eurodesk may not provide the same total support as a central agency would.

Recent developments

As demonstrated in the interviews, secondary school teachers in Aberdeen City and Aberdeenshire relied for European information on the European Commission Representation in Scotland for official EU information, and to a lesser extent, on Eurodesk Therefore, there remained a gap in Scotland in terms of a service that could support the integration of a European dimension within the school curriculum.

In terms of recent developments in this area, the European Resource Centres for Schools and Colleges, initially set up in England and Wales, have since 1999 been extended to include Scotland and Northern Ireland. The only Scottish Centre is hosted by Community Learning Scotland in Edinburgh and shares the same office as Eurodesk. The Centre holds an extensive stock of directories, journals, books and brochures on European issues and has access to a comprehensive database. Its role is to provide teachers and pupils with information on Europe, and give advice on appropriate materials for the introduction of the European dimension by:

- providing information on a particular topic or organisation;
- providing free publications on EU issues in various European languages, many of which are published specifically for schools and colleges;
- offering free access to the Internet;
- acting as a sign-posting service to more relevant services if necessary.

Future plans for the Centre include the creation of a virtual European Resource Centre and in-service training for teachers in Scottish local authorities. These developments are obviously a step forward in meeting the information needs of this sector.

Conclusion

The recent creation of a national network of European Resource Centres for Schools and Colleges is a significant development in the provision of information to the school sector. The results from the survey in Scotland show that there was a need for the network to be extended on a national scale, due to the fact that in Scotland there was a lack of a specific service that could meet the information needs of the school sector. One of the pressing challenges for the Centres will be to provide publications or resources that are designed specifically for and tailored to the needs of teachers and pupils. In addition, the survey has highlighted a number of issues:

- the lack of understanding among teachers of the actual meaning and implications of the European dimension;
- the fact that the European dimension is incorporated in only the modern studies syllabus;
- the information needs of teachers – issues relevant to currency of the information, content and level, and sources and format;
- the lack of European information provision in the school library.

In response to the above, consideration should also be given to professional development issues for both teachers and school librarians in the form of training. At present, the European Information Association [22] is currently running a new course for school librarians and teachers, which examines information available via the Internet for pupils studying the European Union. However, further support is required in all areas that would enable the European dimension to be incorporated into the curriculum in a clear and consistent manner.

Further research is required to examine on a larger scale the European information needs of teachers and, in particular, what they require in terms of resources. In addition, a study of the national network of European Resource Centres for Schools and Colleges would be beneficial in determining its use and success in meeting the needs of this sector.

Although these conclusions are based on a small survey, it would be expected that a similar pattern of results would emerge from additional interviews with teachers and school librarians. Also, there is no data at present to suggest that the national network of European Resources Centres is meeting the information needs of all teachers. This is primarily due to the fact that the Education Relay has only recently been set up in the UK.

Finally, it is necessary to consider broader policy issues and how these are implemented at the local authority level so that the European dimension is incorporated in all schools, and is not left to, or dependent upon, the initiative of individual head teachers of schools or teachers. This is essential to ensure that pupils have the opportunity to learn about other countries, cultures and a wider society.

References

1. *Resolution on the Council and the Ministers of Education meeting within the Council on the European dimension in education of 24 May 1988.* OJ C177, 6 July 1988.

2. *Conference of European Ministers of Education: resolutions: the statement of the Chariman of the CC-PU/Council of Europe, Directorate of Education, Culture & Sport, Higher Education Section*, Vienna, 16–17 October 1991. Strasbourg· Council of Europe, 1991.

3. *A statement of the UK government's policy and report of the activities undertaken to implement the EC resolution of 24 May 1988 on the European dimension in education.* London: DES, 1991.

4. UK Presidency of the European Union Partners in Europe: managing the European dimension: the European dimension in the curriculum. Available at: MACROBUTTON HtmlResAnchor http·//www.dfee.gov.uk/ukpres/partners/manch1.htm.

5. *Scottish education and the European Union: policy, strategy and practice.* Edinburgh: The Scottish Office Education Department, 1994

6. Scottish Consultative Council on the Curriculum. *Thinking European: ideas for integrating a European dimension into the curriculum.* Dundee· SCCC, 1993.

7 Aberdeen City Council Education Department. Policy on the European/international dimension in education, 1997.

8. The European Commission Representation in

the United Kingdom. Available at: http://www.cec.org.uk.

9. Scottish European Relays Network. Available at. http://www.slainte.org.uk/sern/sernhome.htm

10. National Network of Relays targets education. *European Information Service*, 163, October 1995, 30–31.

11. Tackling the information needs of education. *European Information Service*, 167, February 1996, 20–21.

12 New UK Education Relay. *EIA Update* [newsletter of the European Information Association], 56, May 1997, 2.

13. Morton, D. European Union information developments: European Resources Centres for Schools and Colleges. *European Access*, 4, August 1997, 20

14 Morton, D. European Resource Centres for Schools and Colleges. *EIA Review*, 10, July 1997, 14–17.

15. Eurydice: The Information Network on Education in Europe. Available at: http://www.eurydice.org/

16. Central Bureau for Educational Visits and Exchanges. Available at: http://www.british-council.org/cbeve/index.htm.

17. The Representation Office of the European Commission in Scotland. Available at: www.cec.org.uk/offices/sc.htm#THE

18. Eurodesk. Available at: http://www.eurodesk org/

19. European Commission, Directorate-General Education and Culture. *European Documentation Centres*. Available at. http://www.europa.eu.int/comm/dg10/libraries/edc/index_en.htm

20. Cooper, M. Public libraries and EU information. *Aslib Proceedings*, 51(10), Nov-Dec 1999, 346–361.

21. Socrates. http.//www.europa.eu.int/comm/education/socrates/comenius/intro.html

22. European Information Association. Available at: http://www.eia.org.uk/

Visions of the information future[1]

Alison Scammell
Aslib, Staple Hall, Stone House Court, London, EC3A 7PB
alison.scammell@aslib.com

A synthesis of some of the themes and ideas developed in a recently published book about the future of information: *i in the sky: visions of the information future* Common themes included: problems in defining information and defining future time-scales, the ubiquity of information, accessibility, privacy censorship and control, customisation of information products, the development of the World Wide Web, artificial intelligence and cybernetics, changes in working roles and structures of organisations, information literacy, information overload and the organisation and retrieval of information.

Introduction

In November 1999, Aslib published its 'millennium book', a compilation of the personal predictions of the future of information, by more than 40 writers representing a wide range of backgrounds and occupations. As well as information professionals and academics, there are entrepreneurs, journalists, futurists, best-selling management writers, a 'virtual lawyer', a novelist and at least one former rocket scientist. The purpose of the book was to provide a space for writers from diverse backgrounds to be as uninhibited, visionary and as creative as they liked in describing their forecasts. This paper provides a synthesis of some of the book's themes and ideas.

In *The dilbert future*, author Scott Adams declares 'the future is an excellent topic for any author. By the time you realise I was wrong about everything I predicted, I will be dead.' This indicates a major problem associated with writing about the future: it takes courage as well as insight to put your ideas on the line. Many potential authors declined the invitation to write about the future because they did not want to associate themselves with the crystal ball gazing fraternity. Speculating about what might happen in the future was considered a particularly dangerous thing to do, as all too often people are staggeringly wrong in their predictions, and, given the pace of change in the information world, they are unlikely to be dead before their forecasts have had a chance to be proved wrong.

Another problem with writing about the future is defining it. The authors were free to make their own assessments about time scales and, as a result, a variety of projections were included, or, in many cases the exact time span was not specified at all As we all know, the pace of change in information is relentless and events and developments can appear, seemingly overnight, taking everyone by surprise. The advent of the World Wide Web itself left many information providers scrambling to catch up, racing to introduce web based versions of established products. Even Bill Gates was caught on the hop. There is now talk of an 'Internet year' as being less than three months, about the same time the authors were given to write their chapters for the book. With publishing time scales in excess of six months, even for a quick turnaround, the opportunities for expounding predictions that could seem wildly off the mark in twelve months' time, were immense.

---

[1] Based on *i in the sky: visions of the information future*, edited by Alison Scammell, published by Aslib, November 1999. Available (£37 99) from Portland Press Ltd., Commerce Way, Whitehall Industrial Estate, Colchester, CO2 8HP. Telephone 01206 796 351. Email: sales@portlandpress co.uk

What is the future?

For many authors the future is not just imminent and intangible, it is happening already. David Nicholas and Tom Dobrowolski base their ideas around the concept of 'the future is now' and cite visionaries such as Tim Berners-Lee, Marshall McLuhan, Ted Nelson and Vannevar Bush, whose predictions of the future, or observations of the present, directly shaped the progress of events. The other extreme is provided by Richard Wakeford whose discussion of the future extends well beyond the 10,000 year horizon. Wakeford gives the example of the demands of ensuring the safe disposal of high-level radioactive waste (plutonium's half-life of more than 20,000 years means it must be interred for more than a million years), which throws up the question of how to communicate important information beyond the next ice age. The waste store must endure for many times the life of Stonehenge, a structure that after only 4,000 years has become an enigma. The problem here is how to communicate with people who may not be able to recognise current modern languages and who will have forgotten everything about the purpose of the storage structure.

What is information

The other problem facing the authors was how to define information. Is it knowledge or data, is it everything we see and hear, is it smell and touch? David Raitt asks 'does it embrace pretty much everything – remote sensing data, traffic light signals, a sky filled with foreboding cumulonimbus clouds?' The pervasive nature of information is at the heart of one of the major themes of the book, the ubiquity of information. The wholesale integration of information in all aspects of our lives is described by Erik Davis who sees it becoming increasingly:

> more invisible yet more embodied, simply because its flows will literally penetrate the world. Information will no longer lurk in cyberspace, but will be woven throughout our commodities, dwellings, vehicles, even our bodies. Information will be part of the world.

We hear less and less about the information superhighway as information has spread from the perimeter of our working and social activities to become a central force in our mainstream lives.

Accessibility

Information may proliferate but for many it is still inaccessible Kevin Carey's chapter on the future of information and the disabled, discusses the design and customisation issues of information technology from the stance that there is no sharp line between people who are officially defined as disabled and the millions of others who suffer temporary or permanent disadvantage in one way or another. Mike Chivanga also shows how a large proportion of the world's population is denied the advantages of global information systems. In his chapter on the Internet and information revolution in Africa, he is cautious of declaring the Internet as a major force in kick-starting an African renaissance. An estimated 80% of the African population are without access to telephones and electricity and it may take as long as 20 years for the telecommunications infrastructure to improve. Even when information resources are available there is the problem that so much content originates from outside the continent, and is likely to be out of tune with the prevailing socio-economic conditions.

Technology drivers

What will be the technological advances that will contribute to the pervasive nature of information? Throughout the book there are numerous references to intelligent houses and smart household appliances (fridges that can sense when food is past its sell by date and automatically generate shopping lists) and wearable computers (in shoes, ties and necklaces). The possibility of linking humans with computers is also examined. Kevin Warwick's chapter discusses the symbiosis of man and machine and is illustrated with a dramatic example from his own experience. In 1998 a silicon chip transponder was surgically implanted in his left arm, a procedure which was intended to explore the synergy between silicon and carbon, integrating human functions with computer processes. With stories of intelligent pacemakers hitting the headlines and electronic tagging already a well-established fact, this experiment is not as far-fetched as it may at first appear. Indeed, Professor Warwick maintains that before too long thought control of computers will be a realistic development and examples of some recent innovations are included to back up this vision.

Two chapters are written (as fictional futuristic scenarios) showing the likely form of information technology and the information infrastructure of home life. In one, Karl Wiig envisages a range of different IT applications, with different purposes. There will be a FIMS (Family Information Management System), a PIA (Personal Information Assistant) and PWT a (Personal Work Tool). Together these systems provide a holistic solution to information provision. In the other chapter, Peter Bishop introduces the concept of the 'walkabout', a single wireless computing device which meets an individual's complete information needs. It does not crash and remains connected at all times. It can even be trained to sense when its owner wants to make a call and when they do not wish to be disturbed. This, of course, is suggestive of the WAP (Wireless Application Protocol) mobile phones which are being lauded as the next new revolutionary devices, providing 24 hour, world-wide connectivity, devices that will never be turned off and will remain with us at all times, even, we are told, when we are in bed. Apart form the scenario outlined here, there is general agreement that we will have fewer computing devices, probably a single machine, which will cater for all our information and communication needs. The idea of a single computing device is echoed by Ian Brackenbury who predicts that people will not want to buy separate pieces of equipment to print, scan, copy, type, view and store information. Machines will become smaller, more powerful, simpler to use, but above all will be multi-tasking.

Retrieval

The organisation and retrieval of the growing mass of information is, not surprisingly, a common theme. Many authors highlight the increasing importance of portals and other customisable interfaces, developments in knowledge management and advances in artificial intelligence (particularly intelligent searching devices). With the average search engine covering approximately 18% of web pages and the searching effectiveness of all engines accounting for less than 40% coverage of the Web, these developments still seem a very long way off. Gerry McGovern recognises that information overload is such a fundamental problem that it is unlikely to be solved by new search engines alone. McGovern calls for companies and countries to plan for and install a comprehensive information infrastructure, or poor content and poor structure 'will choke off much of the useful information. Poor information will grow like weeds.' Building on the Dublin Core and meta-tagging initiatives provide a start in the right direction, but McGovern advocates systems for categorising information before it is put on the Internet. One way that search agents' performance will be improved, suggests Jack Schofield, is by learning from each other with 'collaborative filtering' techniques. At its simplest these techniques (already in use on sites like Amazon) are based on the idea that people who like one kind of thing will tend to like other similar things.

Ina Fourie argues that although automatic methods to describe information resources will become more important, there will also be greater reliance on what we have learned from thesaurus construction and semantic networks. Fourie also discusses the difficult issue of 'quality' of information, as do so many of the authors, especially with regard to the Internet. Fourie points out the difficulties on deciding what quality really means. The problem with leaving quality assessment to technology and expert systems is that we do not have a satisfactory way of determining quality or an understanding of why human judgement continues to fail.

Overload

Information overload is seen by many of the authors as a continuing problem, unlikely to be solved with advances in retrieval systems.

Gerry McGovern considers information overload to 'be the single greatest problem not simply facing the Internet but facing all of us in practically every aspect of the Digital Age'. Pita Enriques-Harris cites a paper delivered by Noreen Mac Morrow at *Online Information 98*, which advised staying incommunicado as a strategy for dealing with information overload. Enriques-Harris also cites Professor Dover, Chair in Evolutionary Genetics at Leicester University, who foresees a scenario when people will just walk away from the problem of information overload with events such as a 'National Day for Turning Off Machines', or a week when no one can use the phone. One answer to information overload is the need for people to undergo a certain amount of re-wiring of mental processes (which was done before with the transition from the oral to written tradition), a change dependent on a cultural rather than technological revolution, with information literacy skills the key means of accomplishing this. There is a general consensus that the information environment of the future will be user-focused with an emphasis on developing people's information literacy. David Seuss, for example, says he would like his young nephews to grow up to have the same kind of skills as a 'smooth-talking, fast-typing, hacker research librarian. Then their success in the web-based world will be assured.'

A discussion of retrieval and information overload also provides the authors with an opportunity to predict the role of the information professional in the future. Gary Marchionini sees dealing with the growing mass of information as essentially a library problem. The new information specialists will trade as much upon their integrity as their understanding of human information needs and technical skills. Issues of privacy and information overload may cause some people to disconnect from the global information infrastructure altogether. This will lead to a new mediator role for information specialists: 'balancing the information needs of those fully immersed in the global information infrastructure with those who choose not to be, and balancing the need for secure private information with the global hunger for public information'. Information intermediaries will be more like 'self-employed professionals, working in loose coalitions of professionals from a variety of fields who form project-related teams, un-form, and re-form again in different configurations for different projects'.

Information needs

Sheila Webber believes that information professionals will have to find out more about people's information needs and habits. There has not been a paradigm shift in information seeking and the Web has not turned people into super searchers. Information professionals will need to expand their idea of what information is, and to study what individuals do outside the library or information centre. There is a need to explore and create environments that suit people's browsing preferences, to study their habits and how they are motivated to search for, and use information. There needs to be a concentration on studying the cultural implications of information use.

The blurring of information boundaries in our lives will be accompanied by more widespread changes. The authors repeatedly emphasise the relatively minor role technology has to play in the future information age and instead the cultural, political and socio-economic changes are stressed. History has shown (with, for example, the invention of the printing press in the fifteenth century) the highly significant role played by information in revolutionising society. Ian Brackenbury sees a time when there will be a 'complete transformation of entire industries, changes in their relationships with customers, suppliers, and the community in which they operate'.

Organisational structures

Christopher Davis is experienced in transforming organisational structures and re-configuring business models. His firm, Davis & Co, was formed in 1993 around the concept of the use of telecommunications, flexible working and lean management. Information and communication technologies has allowed Davis & Co to adopt the contingent workforce model, with many of its fee earners and support staff being brought in as required. This has provided lawyers with the opportunity to work mainly from home, spending more time with their families, an unusual situation for a profession where people are usually expected to spend long hours in the office. Jack Nilles also writes about the benefits of virtual organisations, which he terms 'evanescent', 'dispersed or distributed organisations...that form and flow, connecting and collecting expertise in response to changes in market/environment conditions'. Frank Colson introduces a new concept integral to the future of both academic and commercial enterprises, that of the 'ateliers'. Loosely defined as:

> studios housing scholars/artisans, engineers and writers working on cognate subjects whose boundaries are defined by no particular technology... concerned with publishing their work in both digital and print form. Small but ever changing in shape, agile yet robust, the atelier will become a primary agent of change in the information society of the future.

This idea of the atelier harks back to the scholar-printers of the fifteenth and sixteenth centuries. Tomorrow's ateliers will herald some far-reaching changes affecting education, artistic and commercial developments along the lines of a new information-driven renaissance. The ateliers will have another fundamental role, concerning authority of documents in the digital era, spurring the development of more powerful multimedia search engines and aiding in the production of high-quality scholarly tools such as thesauri.

Personalisation

The shifting demographics and the rise of the e-generation are shaping the future of the Web and having a profound influence on how e-commerce evolves and encouraging the personalisation and customisation of information products. Don Tapscott describes the shift from a broadcast media to a culture characterised by the new one-to-one interactive marketing model. In parallel to the commercial environment, Tapscott also refers to the paradigm shift in education where the 'one-size-fits-all, lecture-oriented, teacher-focused approach to instruction is being replaced by customised, interactive, student-focused learning which exploits networks and the new media'.

Information will be customised both by the provider and the user, and, of course, this is already taking place. The personalised portal pages of say, Yahoo! or Alta Vista, do not exist until they are looked at, and this has all sorts of implications for storage and archiving of data, and even of establishing a definite version of a piece of material. Information providers are already adept at customising content using cookies, and many authors cite the example of Microsoft's Encarta encyclopaedia which includes different data for different versions: the inventor of the telephone, for example, is given as different people in different language versions. Jack Schofield includes an example of a live Internet performance of Hamlet (as Hamnet) in 1994, a version which brought 'tears to his eyes' but which represented a long tradition of 'customising' Shakespeare, whose plays have been performed in different languages and in various interpretations, for a long time.

Privacy

Concerns about the threats to individual privacy are expressed by several of the authors. Kevin McQueen's chapter on the Internet in China highlights that government's attempt to censor and control the use of the Internet with enforcing registration of web users, arresting 'web dissidents' and blocking websites. Another privacy issue concerns, as David Seuss puts it, 'the real loss of privacy that comes with the Web...the complete transparency that our lives will have to the casual acquaintance, the merely curious, and to the entertainment-seeking-surfer'. The archiving of data sources means that information will remain available indefinitely. This idea is picked up by Barbara Quint who asks us to 'imagine living in a world that cannot forget, that remembers and stores everything – every image, every word, every document, every joke, every insult, every message, every call'. And by Robin Hunt, 'there is no such thing as short-term memory now. Everything is captured. Forever. On a computer.' Clifford Lynch even suggests that in the next centuries we will learn to extract, store and share human memories, with recorded memories becoming part of the record and being managed by cultural heritage institutions. These ideas are given some substance in the last chapter of the book, a short story by Lise Leroux, which takes the capturing of human memories as its central plot.

Conclusion

In re-reading the book it is striking how many paradoxes appear in the authors' predictions. More globalisation of information provision but the rise of niche markets, a glut of data but a dearth of knowledge, information access has the opportunity to provide both liberation and entrapment, connectivity but isolation. How optimistic are the authors in their visions of the information future? Keith Devlin sees the solution to the growing importance of information as developing a common sense understanding of it, based on a firm theoretical (and probably mathematical) framework, which he has dubbed 'infosense'. Devlin claims that without this we will 'handle information with the same dangerous ignorance...with which Marie Curie handled radioactive substances'. Kevin McQueen warns about making assumptions that the information future is so bright and urges us to consider the downsides of technology in the light of the political consequences of use of the Internet in China. Lise Leroux's short story paints a menacing vision of human error in a tale of virtual reality. Many authors refrain from commenting on whether or not their predictions indicate a benign or inauspicious world. For others it is simply too early to tell, or as Erik Davis says 'the jury is out on whether the information age is making us knowledge workers or data slaves'. Perhaps we should bear in mind Charles Handy's comments that although information in the future will be a mixed blessing, we shall adapt and in the end 'life, love and laughter will continue'.

ISSN 0001 253X
Volume 52
Number 8
September 2000

51

Aslib Proceedings

New information perspectives

Comment
When did peer review become anonymous?
A.D. Madden

The London Mayoral websites: cyberdemocracy or cybermediocracy?
Caroline Auty and Alison Cowen

Studying the impacts of the Internet without assuming technological determinism
John A. Daly

Information skills training in support of a joint electronic library in Glasgow: the GAELS Project approach to library courseware development
Nick Joint and Bob Kemp

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

Aslib Proceedings is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In some cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000–8,000 words in length. Articles should be accompanied by a 100–150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The editor is happy to provide advice on an informal basis to potential authors.

Editorial Board

Dr David Nicholas, *Head of Department, Department of Information Science, City University*; John Akeroyd, Director, *Learning Resources, South Bank University*; Sarah Blair, *Head of Publications, Aslib*; Peter Chapman, *Information Manager, NewQuest North East*; Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire*; Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw*; Harry East, *Senior Research Fellow, Department of Information Science, City University*; Mike Hyman, *Managing Director, Head Software International*; Robin Hunt, *Creative Director, AreHause*; Nat Lievesley, *Network Manager, Centre for Policy on Ageing*; Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Senior Lecturer, Department of Information Science, City University*; Richard Withey, *Director, Independent Digital*; Ralph Adam, *Freelance Journalist;* Professor David Ellis, *Department of Information and Library Studies, University of Wales;* Professor Ben Fouche, *Stellenbosch University.*

Editorial Team

Editor: David Nicholas
Deputy Editor: Ian Rowlands
Editorial Manager: Kay Nicholas
Editorial Assistants: Peter Williams (City University), Colette Coles (London Research Centre), Paul Huntington (City University)

CONTENTS

273–276 Comment
When did peer review become anonymous?
A.D. Madden

277–284 The London Mayoral websites: cyberdemocracy or cybermediocracy?
Caroline Auty and Alison Cowen

285–300 Studying the impacts of the Internet without assuming technological determinism
John A. Daly

301–312 Information skills training in support of a joint electronic library in Glasgow: the GAELS Project approach to library courseware development
Nick Joint and Bob Kemp

Comment
When did peer review become anonymous?

A.D. Madden
JFS Learning Resources Centre, 175 Camden Road, London NW1 9HD
admadden@hotmail.com

Anonymous peer review is an important part of the process by which academic information is evaluated. Some of the problems associated with the process are discussed below. Although the history of peer review is well documented, there appears to be no record of when it became anonymous. Since most criticisms associated with the practice result from the anonymity of referees, this is significant. A number of well-known and long-established journals were approached in an attempt to determine when they adopted the process but none of the correspondents at the journals knew.

Introduction

An important part of the process of the evaluation of academic information is the anonymous peer review of articles. The practice is regarded by many as a key to ensuring the quality of academic information. Though few people would argue that there are no faults with the system, most academics accept it with a shrug on the grounds that, it may not be perfect but it is a tried and tested method and has been in place for a long time. But has it? It seems that a surprising amount remains unknown about the origins of a practice on which so much of academia relies.

Faults in the system

In recent years, peer review of research manuscripts for publication:

> has been widely criticised for conservatism, professional prejudice and sloppy standards, with several notorious cases of fraud escaping the notice of peer reviewers altogether [1, p. 3].

Broad and Wade [2] discuss a number of cases of scientific fraud, amongst the best known of which is that of the psychologist, Sir Cyril Burt, whose most famous work was on the inheritance of IQ. The most convincing evidence for Burt's claims came from the study of identical twins that had been raised separately. In 1976, five years after his death, Burt was accused of falsifying his data. Because of his pre-eminence in the field of the measurement of intelligence he was often called on to referee papers which called into question his findings, and was therefore well placed to suppress any work that contradicted him.

Although in recent years some of the evidence against Burt has been questioned, even Hans Eysenck, one of his staunchest supporters, admits that '...the character of Burt was such as to make it possible to believe that he might very well have done what he is alleged to have done' [3, p. 114].

If Burt did abuse the peer review system, as suspected, it was possible because the review was anonymous. Peer review of academic publications has been around for at least three hundred years, since the Royal Society began producing its *Philosophical Transactions* in 1665. One of the justifications for the introduction of anonymity into the review process is that it encourages referees to give unconstrained opinions, and prevents rancorous exchanges between author and reviewer of the kind that famously took place between Isaac Newton and Robert Hooke.[1]

---

[1] Hooke was critical of Newton's papers on optics, and later accused him of plagiarism. In a famous (and uncharacteristic) expression of humility by Newton, he acknowledges his debt to his scientific forebears by claiming that he had seen further than others merely because he had stood 'on the shoulders of giants'. This has been seen as a dig at the conspicuously diminutive Hooke.

The origins of anonymity in peer review

Precisely when anonymity became a requirement of the reviewing process is something of a mystery. As part of a recent research project, an effort was made to answer what was assumed to be a straightforward question, by approaching people associated with long-established journals and asking when their journal had introduced anonymous peer review. Emails were sent to *Nature* and *Science*; but neither Peter Wrobel, Editorial Manager of *Nature*, nor Christine Gilbert, Letters Editor at *Science* (and in her twenty-ninth year with the journal), was able to say when their respective publication had introduced anonymous peer review. The *Journal of Documentation* began in 1945 and has, as far as its Assistant Editor is aware, always used the practice.[2] At the suggestion of the Managing Editor, several long-established members of the Journal's editorial board were contacted, including Graham Mackenzie (formerly librarian at the University of St Andrews), Brian Vickery (Professor Emeritus at UCL), Maurice Line (formerly Director General, British Library, Science Technology and Industry) and Professor Blaise Cronin of Indiana University. None of them knew when peer review became anonymous, and none knew for certain when the *Journal of Documentation* had adopted anonymous peer review; though Professor Vickery was:

> sure that (anonymous) peer reviewing was used at the start of *JDoc* – the first editor, Theodore Besterman, was a meticulous scholar who would have insisted on the practice, which surely goes back a long time before that.

Joanna Corden, archivist at the Royal Society, was not sure whether this was the case or not. The forms that the Society currently uses for referees' reports were introduced in 1898. However, reports exist from as far back as 1832. In 1831, the Statutes, which describe the process by which papers should be judged, were introduced.[3] The Statutes do not say whether or not the reports were confidential.

An approach to Taylor & Francis, publishers of *The Philosophical Magazine* since 1798, led to contact with Jack Meadows, Professor of Information and Library Studies at Loughborough University, who recently co-authored a book [4] on the history of Taylor & Francis. Although Professor Meadows could not give a precise answer to the question of when anonymous peer reviewing became a requirement, he believes that it was some time this century. Peer reviewing, as has been discussed, has been in existence for some time, though the mechanisms for review varied considerably [5]. Professor Meadows suspects that it was the growth in numbers of commercial learned journals following the Second World War that led to the current, relatively standardized process emerging. Anonymity, he believes, is a by-product. Previously, editors uncertain of the appropriateness or merit of a submission would seek the advice of a friend, who might not be known to the author; but this was not usually done on a systematic basis.

Selection of referees

Manheim [6, p. 100] quotes Ziman [7] as stating that:

> The referee is the lynchpin about which the whole business of Science is pivoted. His job is simply to report, as an expert, on the value of a paper submitted to a journal.

The editors of *Nature* choose their referees using a combination of factors, they:[4]

- Consider people who have reviewed or submitted papers to the journal in the past, or who have recently reviewed papers on the same topic (a database of over 50,000 names is maintained).

---

[2] Personal communication, Amber Adams, Assistant Editor, the *Journal of Documentation*.

[3] Those papers that failed to gain a majority vote on two meetings of the Committee were rejected; but prior to the second meeting, the Committee could call upon any Fellow to present a written report to assist the process of deliberation.

[4] Personal communication, Maxine Clarke, Assistant Editor, *Nature*.

- Attempt to choose reviewers from various countries.
- Try to spot conflicts of interest/impartiality.
- Consider reviewers suggested by the authors of submitted papers.
- May consult informal advisers.

Nature was established in November 1869. Its stated objects were:

> first to place before the general public the results of scientific work and discovery and to urge the claims of science to a more general recognition in education and daily life; and secondly to aid scientific men themselves by giving early information of all advances made in any branch of natural knowledge throughout the world and by affording them an opportunity of discovering the various scientific questions which arise from time to time [8, p. 26].

Although the Journal has since become rather too technical for the general public, it has maintained its generality. As such, it is unlikely that it would be susceptible to manipulation by a single, pre-eminent figure, such as Cyril Burt. However, refereeing is a time-consuming and tedious task. A recognized authority in a field, on receiving a manuscript for refereeing, may be too busy to read it and pass it to an inexperienced research assistant for comment. The manuscript could then be refereed by someone with considerably less knowledge of the research area than the author. If the research assistant is acknowledged as the referee, his or her name could then be included on the database, whether or not she or he has the expertise assumed by Ziman [7].

Nature possibly has measures in place to ensure that only reports from nominated referees are accepted: other journals do not.[5] How common this practice is it may be impossible to say, given the secrecy that surrounds the identity of referees. While ideally, more than one referee should be consulted, this is not always possible; and even where it is, an unsatisfactory reference from an inappropriately qualified referee can cause considerable delays to publication.

The origin of academic journals

Magazines began to be published in Europe in the 1660s [9], with the first learned publications (*Journal des Sçavans* and *Philosophical Transactions*) appearing in 1665 in France and England respectively [8].

Houghton notes that:

> Many of the journals which appeared in the seventeenth and early eighteenth centuries foundered after one or two years and in many cases after only one or two issues. A significant reason for their early demise is that they were not underpinned by a sponsoring society... [8, p. 17].

Strictly speaking, neither of the journals mentioned above was officially linked to a 'sponsoring society', when it began. However, the *Philosophical Transactions* was founded by members of the Royal Society and was used to publish the findings and correspondence of the Society's members. It eventually became the Royal Society's official organ on publication of its forty-seventh issue [8]. The *Journal des Sçavans* was never directly linked to any society, but a year after its inception, the Académie Royale des Sciences in Paris was founded [9]. The Académie did not publish proceedings until 1702. Prior to that, reports of the work of its members appeared in the *Journal des Sçavans* [10].

Members of a journal's 'sponsoring society' could therefore use their shared knowledge of the subject matter, and their knowledge of the contributor to evaluate the content of contributions. However, as the readership of journals began to grow, it rapidly became impossible for all the members of the community associated with a given journal to know each other. The appointment of editorial boards and referees amounted to the selection of representatives of different sections of the community. This view of editors and referees as representatives is reflected in Ziman's remark that:

> ...the publication of scientific papers is by no means unconstrained. An article

---

[5] As a newly employed post-doctoral researcher the author had a number of papers passed to him for refereeing. Although qualified to respond, he was largely unpublished.

in a reputable journal does not merely represent the opinions of its author; it bears the *imprimatur* of scientific authenticity, as given to it by the editor and the referees he may have consulted [7, p. 148; 11].

Where those referees are anonymous, however, it is impossible to know how representative they are. An important component of the natural evaluation process is therefore missing.

Alternative approaches

Because of criticisms associated with anonymity, some journals have chosen to make known the identities of their referees [6, p. 101]. This, not surprisingly, reintroduces the problems that anonymity was intended to overcome, not the least of which is finding volunteers willing to act as referees, since they now not only have to perform a tedious task, but they risk provoking animosity.

If Professor Meadows is correct, and anonymous peer review was introduced after World War II, then it seems likely that alternatives were considered and rejected. It would be fascinating to have an insight into the associated discussions; but despite the fact that the practice appears to have arisen within living memory, no one seems able to remember when.

References

1. Grayson G. *Scientific deception: an overview and guide to the literature of misconduct and fraud in scientific research.* London. The British Library, 1995

2. Broad, W. and Wade, N. *Betrayers of the truth.* New York: Simon and Schuster, 1983.

3. Eysenck, H.J. Burt as hero and anti-hero: a Greek Tragedy. Chapter 6. In: Mackintosh, N.J., ed. *Cyril Burt: fraud or framed.* Oxford: OUP, 1995.

4 Brock, W.H. and Meadows, A J. *The lamp of learning.* London: Taylor & Francis, 1998.

5. Meadows, A.J. The problem of refereeing. In: Meadows, A.J., ed. *The Scientific Journal.* London: Aslib, 1979, 104–111.

6. Manheim, F.T. The scientific referee. In: Meadows, A.J, ed *The Scientific Journal.* London: Aslib, 1979, 94–103.

7. Ziman, J.M. *Public knowledge: the social dimension of science.* Cambridge: CUP, 1968.

8. Houghton, B. *Scientific periodicals: their historical development, characteristics and control.* London: Clive Bingley, 1975.

9. History of publishing. In: *Britannica CD 2000 Deluxe*. Encyclopaedia Britannica Inc, 1994–2000.

10. McKie, D. The scientific periodical from 1665 to 1798. In: Meadows, A.J., ed. *The Scientific Journal.* London· Aslib, 1979, 7–17.

11 Zuckerman, H. and Merton, R.K. Patterns of evaluation in science· institutionalisation, structure and functions of the referee system. In: Meadows, A.J, ed. *The Scientific Journal.* London: Aslib, 1979, 112–146.

The London Mayoral websites: cyberdemocracy or cybermediocracy?

Caroline Auty and Alison Cowen
Library, House of Lords
autyc@parliament.uk
CowenA@parliament.uk

The London Mayoral Election of 4 May 2000 was probably the first in British electoral history when all the candidates had established a presence in cyberspace. An analysis of the 11 websites showed that politicians are beginning to realise the potential influence of the Web in political circles, both as a means of publication and also for canvassing support. Most of the web pages were well designed with a reasonable amount of content and simple levels of interactivity. Multimedia features were also evident on most sites, though in terms of currency and timeliness of information, some improvements could still be made.

Introduction

Internet use by political parties and politicians is certainly a hot topic. If the penultimate General Election was dominated by pagers and mobile phones [1] then surely the forthcoming one will be conducted at least partially in cyberspace. However, a review of political party websites after the last General Election [2] showed that although most parties had established a presence on the Net, few (and only the smaller parties) had capitalised on the real benefits of politics on the Internet, namely the use of multimedia features, real time updating of information and two way interaction. Perhaps the elections for Mayor of London would be the turning point; after all London must be one of the most connected places in the country with '74% of companies making frequent use of external networking technologies' [3], not to mention people who log on at home. Certainly, the websites of the candidates have all received coverage in the press [4] and featured on the BBC, *Guardian* and *Evening Standard* portals. Similarly, the candidates from the main political parties (Dobson, Norris and Kramer) all took part in cyberdebates organised by *The Guardian* at guardianunlimited.co.uk/mayor – so have people finally realised the potential the Web has to offer? We shall see.

Methodology

Sample

The nominations for candidates standing for Mayor closed on 3 April and the official list was released by the Returning Officer on 5 April. Despite claims that Winnie the Pig in Parliament Square was standing on a 'Support British Pigmeat' [5] ticket, the list consisted of only 11 candidates. The BBC website [6] immediately mounted links to those candidates with websites. Originally 10 out of the 11 candidates supplied their URLs (Michael Newland of the BNP refused to furnish the Corporation with any details of his campaign), but eventually all candidates established a website for the campaign, probably the first time in British electoral history that this has happened.

The candidates for Mayor were:

- Geoffrey Ben-Nathan, Pro-Motorist and Small Shop
  http://www.pro-mass.co.uk/
- Geoffrey Clements, Natural Law Party
  http://www.the-mayor.org.uk
- Frank Dobson, The Labour Party
  http://www.frank-dobson.org.uk
- Ram Gidoomal, Christian Peoples Alliance
  http://www.cpalliance.net/
- Damian Hockney, UK Independence Party
  http://www.independence.demon.co.uk/html/london_mayor.html

- Darren Johnson, Green Party
  http://www.greenparty.org.uk/GLA2000
- Susan Kramer, Liberal Democrat
  http://www.susankramer.org/
- Ken Livingstone, Independent
  http://www.livingstoneforlondon.org.uk/
- Michael Newland, British National Party
  http://www.bnp.to/gla/
- Steve Norris, Conservative Party
  http://www.norrisforlondon.com/
- Ashwinkumar Tanna, Independent
  http.//www.relaxconsultancy.com/ashwin.htm [7]

The evaluation criteria

The criteria consisted, broadly, of examining the content, design, currency, interactivity, external links, and multimedia features of the sites. The evaluation period ran from 5 April (when the list of candidates was released) to the week after the elections for Mayor on 4 May 2000. The criteria were applied as a whole initially and the sites were then checked weekly to see which features had been updated and the overall currency of the sites. As seen from the URLs above, most of the candidates established an independent presence in cyberspace, independent from the party sponsoring them, that is. However, of those few (UKIP, BNP, Green Party and CPA) whose websites were part of a party site, only the pages relevant to the London elections were included and analysed.

Results

Design of the mayoral websites, as with all websites, was felt to be a critical factor in the evaluation framework, as candidates had to appeal to a wide audience of voters, with different levels of experience of the Net. At the very least each site should provide a menu bar, navigational tools and a helpful colour scheme and appropriate font size. Moreover, attention should be paid to metadata – candidates making full use of this facility had clearly got to grips with more advanced features of HTML. In the main, the home pages of the 11 sites followed a similar style. All of the candidates provided menu bars and navigational tags (back and forward arrows and home tags), though a number of the menus did prove a little unhelpful. The ProMaSS party menu consisted of 'Home, Says, Policies, Hypocrisy and Feedback'. Although striking, the terms used did little to aid navigation around the site. Other candidates suffered from poor categorisation of their site. For instance, Ram Gidoomal had 'News' as a category in the menu bar and a link to 'Press Releases' and 'London News' on the main page, and no definitions or explanation to help the user.

Ram Gidoomal of the Christian Peoples Alliance offered one of the most striking home pages. The page consisted of a simple white background and carried only a photograph of Mr Gidoomal with his name and that of his party. An applet ran underneath the photograph, asking 'Could this man be the next Mayor of London?' The purple writing against the simple white background was very effective, particularly as it made a contrast to the more cluttered sites of the other candidates, such as Susan Kramer's site. Similarly, the BNP site opened with a flashy shockwave banner, which then gave way to the main site. The candidates from the main political parties used the traditional colours of their organisation on their home pages and consistently throughout their site. As an independent candidate Ken Livingstone employed the colour purple. On the home page it screamed 'Paint London Purple' and encouraged people to display posters and banners. The consistent use of purple on the site and throughout the campaign proved to be an extremely effective tool.

Nine of the eleven mayoral sites had included some metadata. The BNP and Ashwin Tanna's site did not have any keywords to aid retrieval. Noticeably, Steven Norris included the names of some of the other candidates in the meta name 'keywords' field. He had the names of Frank Dobson and Ken Livingston, but there was no mention of Susan Kramer. By contrast, Susan Kramer had only named Ken Livingstone, suggesting that she saw him as the main rival. Tellingly, Ken Livingstone incorporated the names of celebrities like Chris Evans, Seumas Milne and newspapers such as *The Guardian*. Perhaps the most striking use of keywords was by the ProMaSS party who even included the term 'yellow lines' and 'wallets on wheels'! None of the websites included all 11 of the names of the candidates and few used obvious terms such as the GLA or elections.

Overall, most of the 11 sites had a consistent/uniform look and feel to them, suggesting they had been purposefully designed rather than haphazardly thrown together. A few candidates would have benefited from a few lessons in the finer points of design: Geoffrey Clements could have sectioned his home page up into more manageable segments and Ashwin Tanna could have used fewer colours in his colour scheme (almost every page had a different colour background). Damian Hockney perhaps could have aligned the menu buttons with the relevant text menu. Out of the 11 candidates only the Green Party (Darren Johnson) made an obvious attempt to increase the accessibility of its site. Advice was given on how to increase the font size for the partially sited, to decrease the font size for use with older monitors and how to select the 'no frame' version of the site.

Content

The design of a website may provide the skeletal structure but the substance comes from the content. In terms of the type of information found on the Mayoral sites, there were many similarities: biographical details, press releases, policies, manifestos and campaign materials featured on most sites. Some candidates were more imaginative and broadened the scope of their pages, immediately setting them apart from their rivals in cyberspace.

In terms of the policy fields covered, the GLA and Mayor will have responsibility for eight distinct areas:

- Transport
- Planning
- Economic development and regeneration
- The environment
- Fire and emergency planning
- Culture
- Health
- The Police [8]

So at the very least, all the candidates should have had content covering these topics. Up until the week beginning 21 April 2000 (two weeks before the election) Ken Livingstone's policy page consisted solely of the phrase 'this page will be expanded over the coming weeks to give a comprehensive guide to Ken's policies for London'. Not very helpful and something that was picked up in the press [9]. However, the launch of his manifesto redeemed the situation, most of the delegated areas being represented with sections on 'improving the health of Londoners', 'a world capital for arts, culture and sport', 'new prosperity and opportunities for Londoners' and 'getting London moving' to name but a few. In fact, all of the mainstream candidates had developed policies covering the eight designated areas – Susan Kramer's manifesto was divided into three main sections: 'rescuing public transport', 'reducing crime' and 'revitalising communities' with subsections on 'working for a healthy London' and 'a world city – a city of many worlds'.

Both Steve Norris and Frank Dobson had the key eight points covered. Candidates without the backing of a major party (Ken Livingstone excluded) tended to stray from this model. Geoffrey Ben-Nathan of the Pro-Motorist and Small Shop party had two principal tenets in his policy section: 'Londoners' right to drive' and 'Londoners' right to park'. The whole section on policies covered one page (Susan Kramer's manifesto alone was 15 pages, Frank Dobson's 32) three-quarters of which concerned roads. Hardly surprising for a single-issue candidate but there was a handy get-out clause for the lack of other policies at the end of the section: 'Many things in London work well – without a Mayor. A ProMaSS Mayor would therefore be dedicated to not interfering.' In a similar vein, Damian Hockney's manifesto for the UK Independence Party started with 'London needs its £2 billion back from Brussels, and a global future, not an impotent Mayor and Assembly run by the control freaks of Millbank and Brussels.' The entire document was a tract against the EU, the only mention the devolved areas received was in the sentence

'London must lead in turning back the tide of EU regulation and must bring back the money to spend on health, wealth creation, transport, infrastructure and housing.'

Geoffrey Clements of the Natural Law Party appeared to have touched upon many of the delegated areas (health, transport, employment, law and order) in his policies section, but also extended the brief of the Mayor to include making London 'capital of world peace' and a 'healthy food – GM-free zone' Clements also claimed that the Natural Law manifesto was a 'blue print to create Heaven and Earth in the city' and the solution to the Northern Ireland crisis (clearly not one of the Mayoral responsibilities) lay in reducing stress levels. So, although some of the candidates had a clear understanding of the responsibilities the Mayor will have and have developed policies accordingly, others did not really focus on the issues.

Although one might have expected there to be considerable overlap in the areas covered by candidates, this proved not to be generally the case. However, in terms of other content on the web pages, there were some similarities. Most of the candidates had biographical details on their sites (even the BNP). All had press releases and speeches, though some candidates had made greater efforts in this area than others. Ashwin Tanna (Independent) had only four press releases covering his campaign, whereas Ken Livingstone's press archive stretched back to October 1999 and contained almost fifty releases. Most candidates had a section on how to help their campaign by donating money (see section on interactivity) or time. The more mainstream candidates had copies of articles in the press about the campaign, mainly from the broadsheet press, though Mr Tanna's articles came from the *South London Press* and *Chemist and Druggist*, not having received any mention in the national press, one assumes. With so much common ground in terms of content, what really caught one's attention were those candidates who had made efforts to make their sites unique and a few deserve special mention here.

Most people had policy details in a simple list with subject headings – not really making full use of the hypertext features of the Web. Frank Dobson's site was more like a gateway – his policies were listed in a menu and if you selected a topic (such as crime) a result list was displayed that included policy statements, speeches, press releases and other articles.

Ken Livingstone had a whole section devoted to the Labour Party selection process that included statements issued by him at various stages, an open letter to Tony Blair, and supportive excerpts from the press – all in reasoned language to justify his action of standing as an Independent. Both Livingstone and Dobson had lists of Labour MPs or celebrities who supported their cause. Another, Ashwin Tanna, had a section containing letters between himself and the BBC about the lack of election coverage for minority candidates.

Both Frank Dobson and Susan Kramer had campaign diaries to chart and report their every move. Both deserve credit for this, but Kramer's was even more imaginative as it was written (as if) by her cat Whittington under the title 'Paws for Thought'. Susan Kramer also intended to visit every high street in London (174 in total) with her campaign entourage and there was a section on her website that showed dates and times of forthcoming visits. The UKIP also had an events page but this only had two entries: the first saying when polling day was, the second when the nominations for the UKIP mayoral candidature closed.

Since this was the first election of its kind, it is somewhat surprising that so few of the websites had information on the functions of the Mayor and Assembly, the electoral process and how one might use one's second vote. Damian Hockney of the UKIP ventured into these areas with a series of fact sheets on London, a summary of the White Paper and information on the elections for the GLA. Michael Newland (the BNP candidate) also had a page entitled 'How it will all work', which explained the voting system to be used. Darren Johnson of the Green Party and Ram Gidoomal also deserve praise for discussing the London Assembly and the GLA structure. Only Ken Livingstone suggested how electors might use their other preferences in a section entitled 'Ken recommends Labour and Green vote for Assembly':

Ken Livingstone for Mayor
Labour in the single-member Assembly seats
Green in the London-wide top up lists
Give second preference votes to Mayor to Frank Dobson

Other candidates seemed to assume that having external links to the DETR site on London governance or White Paper excused them from explaining the system.

Interactivity

Interactivity is one of the key features of the Internet as it provides opportunities for real time feedback for policies in the form of emails or even online opinion polls – two way communication instead of the usual top down approach of most forms of publishing. Online forms can also be used to boost party support and finances. In this area, it seems that the Mayoral candidates have realised the benefits to be gained from opening up as many channels of communication with the electorate – something for which they all deserve credit.

At the very simplest level, interactivity exists in the form of mailtos. Almost without exception the web pages we examined included either email addresses or clickable mailtos. Ken Livingstone being the exception to the rule. For Damian Hockney of the UKIP this was the sole extent of interactivity on his web pages, while Geoffrey Clements had an email address but not in clickable form – missing the boat slightly. Others had made greater efforts: Frank Dobson had mailtos for most of his support team, including the press office. Dobson also offered an email news service whereby people could register to receive updates and new information via email. He was not alone in this – Steve Norris, Ashwin Tanna and Ken Livingstone also offered this, though there was a problem registering for the latter. The BNP also had this facility on their web page though it was not specific to the Mayoral elections and thus was discounted. How many emails were sent and on what subjects is another (interesting) matter though.

Five of the six sites that offered email updates were sent an email address. [It was impossible to register with Ken Livingstone's.] Frank Dobson's site was particularly impressive, as the user had to subscribe to the service in the same way that you would sign up to a discussion group. Almost as soon as the contact email address was sent, a reply was received which explained how to join and how to unsubscribe from the service. Details about how to contact the Frank Dobson list owner were also found in the preliminary message. In all, twenty-five updates were received after having subscribed on 7 April 2000. The up-dates concerned numerous subjects such as Frank Dobson's policy on crime, his ideas on job creation and included a statement given by Tony Blair, urging people to 'turn to Frank'. A number of the articles sent were duplicated which was a little annoying and many of them attacked Ken Livingstone and his policies, with one denouncing Ken Livingstone as 'unfit to be the Mayor of London' After 3 May no more up-dates were received. Disappointingly, only one other site replied to the email. Ashwin Tanna's website sent a reply to the initial email which stated 'Thanks for your interest. You have been added to the mailing list' but no further messages were received.

Both Frank Dobson and Ashwin Tanna had email addresses specifically for questions to be mailed, which they then answered and posted the replies on the Net. Unfortunately, well into April Mr Tanna had not received any questions to respond to – not for want of trying though. Many candidates had provided online forms for credit card donations to support the campaign – Steve Norris had this along with Michael Newland of the BNP and both deserve praise for having secure servers (https in the URL and padlock at the bottom of the screen). Electronic commerce is to be encouraged, but could candidates really expect people to send their credit card details on unsecured servers – as Geoffrey Clements request for £50 appeared to be. Ken Livingstone also requested financial support, but asked you to print out the form and post it to him – hardly making full use of Internet technology, but better than running an insecure server His form for purchasing campaign merchandise also required the use of snail mail.

The best examples of interaction on the Net are opinion polls. Only a few candidates ventured into this territory. Livingstone had a link to the vote.co.uk page asking you to vote for him in their online poll. Darren Johnson and Ram Gidoomal also linked to vote.co.uk. The latter candidate asked people to register their intention to vote for Ram Gidoomal in advance of the May elections to 'show London that Ram is a serious contender'. Frank Dobson's page opened with a question asking how you would spend £50 million in London – the choices were health, education and transport. When you made your choice it showed you how other connected electors had voted. Sadly, the question remained the same for the duration of the campaign. The BNP claimed to have a daily poll – the one for 28 April asked 'Do you think the homosexual lobby within the Labour Party is influencing the push to abolish section 28 leading to homosexual propaganda being foistered on school children against the wishes of the public?' This was illustrated with a picture of Peter Mandelson at the dispatch box. You could click to see the results of this survey.

Multimedia

This is another key feature of the Internet, which sets it apart from other forms of publishing – the ability to mix sound and moving images. So were the Mayoral sites multimedia savvy or multi-mediocre? To begin with, the sites were lacking in this area but as the campaign progressed, sounds and video clips were added and thus enriched the web pages. The only site that did not embrace multimedia was the ProMaSS website. However, even the simplest of sites had some multimedia features: Ashwin Tanna's page on crime was illustrated by a burglar carrying a swag bag – his legs moved as if fleeing the scene of a crime. Many of the sites opened with an audio message – Ken Livingstone's had this, as did Dobson's and Geoffrey Clements' (Natural Law). Susan Kramer went one better by having a real audio introduction to the site and a version of her theme tune 'Follow my Footsteps'. Michael Newland's pages had a shockwave opening and links to three pages of video clips. Although some of these clips were by other Mayoral players, there was nothing to suggest that the clips had anything to do with London or the GLA elections. Frank Dobson had both audio and video excerpts that did appear to be related to the campaign – the audio files included 'Frank outlines his plan for London Underground' and 'Frank introduces his transport policies for London', while the videos included one of his party political broadcasts. He was also the only candidate to have his own screen saver available for down loading, which showed a rotating panoramic shot of the London skyline. Although multimedia features have the potential to exclude voters reliant on older technology, overall the efforts made in this field were praiseworthy and showed that political players are finally embracing the true nature of the Net.

Images

Images are a simple way of enriching a website and one which all candidates made use of, usually by having a picture of themselves in the top left hand corner of each page. Ashwin Tanna not only had this but also a gif of Tower Bridge on his home page. The award for most imaginative use of images went jointly to Susan Kramer and Frank Dobson. Susan Kramer had a picture of her close family, Whittington and the dog (Molly) on her site – thereby appealing to pet lovers everywhere. Dobson had a whole page of images of himself and his Deputy, Trevor Phillips, in various poses and even showed the byte sizes, should you wish to copy the images. He also had wallpaper to download and banners to insert into your own web page – with full instructions and HTML coding provided. All very creditable. Steven Norris's website had also made full use of images. He had eight photographs, two of himself and his wife, three of a trip to Bangla Town and the remaining three showed Steven Norris enjoying a day out at Wembley. It was possible to view the photographs in 320 × 240 or 640 × 480 format and the byte size of each had also been provided.

External links

The original intention in this part of the evaluation was to examine whether candidates maintained their web pages and updated out of date links. It soon became apparent that looking at where a potential Mayor linked to was much more revealing. However, some candidates had clearly missed the boat here. Ashwin Tanna had no external links and Geoffrey Clements only one – to the main site of the Natural Law Party. The ProMaSS Party simply linked to the company that created their site. In a similar vein, Michael Newland's site was very inward looking and only linked to the main BNP web page, the Young BNP and America BNP. Both Livingstone and Dobson had made greater efforts in this area, and their choice of links was quite similar. Livingstone linked to the pages of some of his rivals – Dobson, Kramer, Johnson and Norris. Dobson reciprocated these and included a link to Ram Gidoomal of the Christian Peoples Alliance. Why neither of them had links to all the candidates and had picked only those few is unclear. Both had links to external bodies concerned with transport, crime and London in general. Both had links to various media organisations, though Livingstone only linked to the *Evening Standard* and *New Musical Express* – Dobson had links not only to printed media but radio stations and web portals covering the Mayoral campaign (ITN the Race for London Mayor and News Unlimited A Mayor for London). Frank Dobson had a section devoted to London governance which included a copy of the original White Paper, the results of the referendum, areas covered by the GLA and information on the electoral system, all hosted on external sites. The links to the Labour Party (local and national) and related bodies (Fabian Society, trade unions) found on Dobson's page were (not surprisingly) absent from Ken Livingstone's, though he did represent the environment, anti-racism and students on his page of external links (ignored by all other candidates). Steven Norris had grouped his links under three main headings, 'other candidates', 'sites with sections on the campaign' and 'newspapers and news organisations'. He listed the Conservative Party, tubehell.com, vote.co.uk and Londonelection.org under the campaign section.

Currency

A large percentage of most of the candidates' websites could be left once they had been created but press releases, diaries and events pages need to be updated to be effective. So, how timely was the information mounted? Even after the elections on 4 May, the UKIP site helpfully told readers 'It is expected that the first elections to the Greater London Authority will be held on 4 May 2000.' Conversely, on polling day Steve Norris's site stated 'By 9 o'clock tonight, Londoners will have made a decision they have to live with for 4 years....' Similarly, even before the results for Mayor had been released, Frank Dobson's site had been dismantled except for a statement which read '...I hope that the new Mayor will harness the great opportunity they have in tackling this gross inequality and put their heart and soul into reversing the ever-growing gap between rich and poor.' None of the remaining candidates marked election day in any way. In terms of currency throughout the campaign, however, some candidates fared better than others did. On 17 April, Dobson and Livingstone had press releases on their websites that had been issued that very morning. Compare this with Susan Kramer who was a whole week behind and Steve Norris, whose latest news was dated 28 March. On 28 April (one week until polling day), Dobson's most recent press release was dated 25 April, though one of the cuttings from the press was from the Evening Standard of 28 April. Darren Johnson of the Green Party had the most up to date item (dated 28 April) with Livingstone trailing behind on 26 April and Kramer further back on 24 April. Steve Norris still lagged behind – his most current notice was dated 20 April.

Conclusions

In some senses the Mayoral campaign conducted on the Internet mirrored the nature of the real one – Frank Dobson's site had to be redesigned and relaunched after it was hacked [10] and a spoof Dobson site was set up at www.ooohyouamazeme.com.

However, this was surely the first British election where all the candidates established a website and, on the whole, the sites were a marked improvement on previous political incursions into cyberspace. Most of the sites had a consistent and corporate look and feel to them and, without exception, navigational aids were provided. The use of metadata by the majority of the candidates suggests that use of HTML has progressed beyond the obvious. In terms of content, there was much common ground – most of the Mayoral hopefuls stuck to the traditional format of manifestos, press releases, biographical information and cuttings from the press. A few should be praised for being more imaginative and having items like diaries (feline or otherwise) or downloadable campaign posters. In terms of interactivity, with the exception of Ken Livingstone, mailtos or clickable emails were available on all pages. Most candidates presented electronic forms to enable surfers to offer support or funds or even question those standing (though not all of the transactions were secure). Online opinion polls were present only on a minority of sites, suggesting that this degree of cyberdemocracy still has some way to go, though the number of email update services was encouraging. Multimedia features were in existence on almost all the sites. Overall then, less a case of cybermediocracy and more the green shoots of emerging cyberdemocracy.

References

1. Jenkins, M. The wired election? Network, *Independent*, 25 March 1997, 2.

2. Auty, C and Nicholas, D. British political parties and their web pages, *Aslib Proceedings*, 50(10), 1999, 283–286.

3. DTI. Moving into the Information Age. a regional benchmarking study, 1999. http://www.isi.gov.uk/isi/bench/regional99.htm (visited 10 May 2000).

4. Holliday, R. and Leapman, B. Loudhailers out, slick websites in. *Evening Standard*, 19 April 2000, 23.

5. Press Association. http://www.pa.press.net/news/story/Animals)London-Mayor-pig_139554.html (visited 1 April 2000).

6 BBC. http://news.bbc.co.uk/hi/english/in_depth/uk_politics/2000/london_mayor/default.stm (visited 5 April 2000).

7. Greater London Returning Office. *Election address by candidates for Mayor of London*, April 2000 – websites added later.

8. DETR. *London governance fact sheets: Greater London Authority*. http://www.detr.gov.uk/London/fact sheets/001.htm (visited 12 April 2000).

9. Cllr Jake Turnbull, letter, *The Guardian*, 22 April 2000, 17.

10. Guardian Diary, *The Guardian*, 21 April 2000. http://www.guardianunlimited.co.uk/Archive/Article/0,4273,4010422,00.html (visited 21 April 2000).

Studying the impacts of the Internet without assuming technological determinism[1]

John A. Daly
14205 Bauer Drive, Rockville, MD 20853-2339, USA
daly@erols.com

Developing countries generally have low levels of Internet services, and as a result require different approaches to impact measurement than developed countries. A conceptual framework is proposed which accepts a role for technological innovation, but which rejects technological determinism. It recognizes that economic, social, political and cultural factors affect the penetration and use of the Internet. It emphasizes direct and indirect impacts of the Internet on people, while including impacts on institutions and the environmental factors and policies that affect institutional impacts. Ultimately the Internet is an induced innovation, but developing countries still suffer from the Matthew principle – that those who have most will be given still more.

Impacts of the Internet range from communications cost savings, to changes in performance of individual businesses, NGOs, government agencies, and schools, to changes in performance of markets, to those measured in terms of economic growth, equity, health status, knowledge, and environmental quality. The overall view of the impacts of the Internet emerges (as does the picture in a jigsaw puzzle) from combining many studies of specific effects, each conveying a part of the picture.

The majority of the people of the world live in low and middle income countries; they have the greatest need of the Internet to help solve the pressing problems of poverty, and they are the least prepared to use the technology and appropriate its benefits. Clearly great benefits are available to developing countries from appropriate uses of even their scarce Internet networks.

Many of the institutions affected by the Internet are international. Internet impacts on these international institutions must be confronted. Thus developing countries may face significant risks from participation in international financial, labor and goods and services markets, because of significant gaps in connectivity and in knowledge and information. On the other hand, they may benefit greatly from power uses of the Internet abroad, of which Africans are scarcely aware, such as famine early warning and epidemiological alert systems.

Donor agencies encourage the development of the Internet in developing countries, and especially in Africa. Several have agreed to work collaboratively to learn the lessons from their experience. Such efforts are important if the potential of the Internet is to be realized in developing countries, the risks inherent in the Internet are to be avoided or ameliorated, and the net effect to be enhanced equity and social and economic growth.

---

[1] This article is taken from Nicholas, D. and Rowlands, I., editors. *The Internet: its impact and evaluation. Proceedings of a forum held at Cumberland Lodge, Windsor Park, 16-18th July 1999.* Aslib, 2000.

Introduction

In 1746, at the grand convent of the Carthusians in Paris, Jean-Antoine Nollet placed 200 monks in line, each linked to the next by a 25-foot long iron wire. Conducting the world's first test of electrical signal propagation in a network, Nollet connected the chain to a battery.

> The simultaneous exclamations and contortions of a mile-long line of monks revealed that electricity could be transmitted over a great distance; and as far as Nollet could tell, it covered that distance instantly [1].

Would that all impacts of networking were so visible, and few so painful.

The theme of this paper is the 'impact of the Internet'. It is a theme that reeks of technological determinism, suggesting that the networking technologies and the underlying information and communications technologies on which they rely are in themselves affecting society. Rapid improvements in these technologies have been related to the rapid increase in their use. Indeed, technological determinists might suggest:

> Information technology is like electricity in its ability to transform the organization. Unlike electricity, however, the latest technological boom has spread wide across industries and penetrated deeply into corporations in a relatively short period of time [2].

The position of Marx and Smith seems more relevant:

> In spite of the existence of an engineering profession, technology is not an organized institution; it has no members or stated positions; nor does it initiate actions. How can we reasonably think of this abstract, disembodied, quasi-metaphysical entity, or one of its artifactual stand-ins (e.g. the computer), as the initiator of actions capable of controlling human destiny [3]?

The penetration, utilization, and impacts of the technology depend on social and economic factors, and on the actions taken within countries relevant to the Internet.

What impacts are of interest?[2]

One of the better studies of the Internet in Africa focuses on the cost savings achieved by African users who substitute email for telephone calls and faxes [5]. Surely such consequences of use of the Internet are significant, and should be considered as 'impacts'. A review of the literature on Internet impacts quickly shows an interest in the consequences in terms of the objectives of many organizations: schools, legal offices, health facilities, private voluntary organizations, governmental agencies, manufacturing firms, financial service firms, etc. We too must consider such consequences as 'impacts'.

Further, I have interviewed senior managers in donor organizations, and they ask, 'What are the developmental consequences of the Internet?' They must consider the value of investments in the Internet versus investments in other technologies or programs in terms of reduction of poverty, improvement of health, improvement of education, economic growth, improving citizen participation in governance, promoting the rule of law, and preservation of the environment. But, in Africa, with more than 99.9% of the population unconnected to the Internet, how are we to trace Internet effects on such broad developmental concerns?

These interviews indicated more interest in the impacts of policies than in the impacts of the Internet *per se*. What policies and policy instruments can one use to make the impacts of the Internet more equitable? To help developing countries utilize the Internet more fully and effectively to promote economic growth? Thus, one is faced with the difficult task of determining not only the consequences of the introduction of the Internet, but also how those differ from the developments that would have occurred under other circumstances.

A conceptual framework

There will be many studies of the impacts of the Internet. A framework is required to suggest how those studies may be interrelated, and indeed how one may extrapolate from relatively narrow studies to understand more general phenomena. Thus, one might expect many studies of the effects of the introduction of the Internet in businesses, and some studies of the effects of the introduction of the Internet in markets (to improve dissemination of information among buyers and sellers). Understanding of the penetration and use of the Internet in a region, combined with understanding of the effects of use and non-use of the technology in firms and in markets, should contribute to the understanding of the effects of the Internet on, say, the small-and-medium-enterprise sector or on a specific industry.

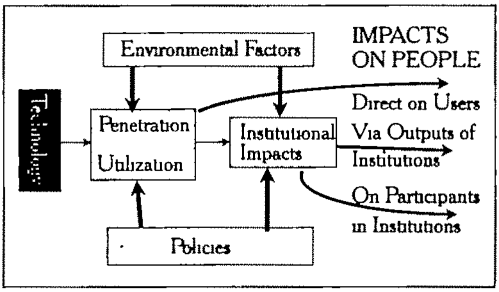
---

[2] Unfortunately the definition of Internet is not consistent from author to author [4]. The following discussion is based on a relatively conservative definition of the Internet.

In this sense the framework may be considered to provide a 'jigsaw puzzle' approach, in which a picture of the large scale impacts of the Internet emerges from piecing together many small studies, each providing understanding of a portion of the whole.

The difficulty of understanding the impacts of the Internet should not be underestimated. For example, a case-control study of small enterprises using and not using the Internet would be likely to show considerable differences in effects on companies within each group – perhaps better illustrating the range of effects available to the enterprises rather than any single 'impact'. Moreover, the difference between groups would be illustrative of the differential impacts of the Internet on users versus non-users. If one were looking at booksellers in Africa who were or were not going online, it might well be difficult, if not impossible, to identify the impact of Internet booksellers like Amazon.com as they affected the African market, or the effect of authors and publishers using the Internet to improve the quality or lower the prices of their products, and thus affecting the goods sold by the booksellers. Similarly, it might be difficult or impossible to identify the impact of the Internet in opening other information choices for the clientele of the booksellers, as an impact on their market. The best chance to understand such complex impacts is to step back and look for larger patterns among many diverse studies.

Figure 1 *Open-loop model of impacts of the Internet*



Drawing on the framework put forth in a US National Academy of Sciences publication [6], a reductionist approach is proposed. It is illustrated in Figure 1.

The framework considers the penetration of the Internet and its utilization by those connected. The penetration and utilization are seen as affected by technological forces, as well as by policies and 'environmental' forces outside of the control of policy makers. Similarly, the framework considers institutions and impacts of the Internet on institutions The framework recognizes that the institutions are inter-linked in complex patterns. Impacts on people may be either direct, stemming from their 'consumer' uses of the Internet, or indirect arising as a result of their participation in institutions, or as a result of changes in the ways institutions serve their clients.

Technological change as a driving force

Technological change clearly underlies the development and growth of the Internet. Moore's Law states that the processing power of the transistor chip doubles every 18 months, (and thus technological change drives costs down);[3] Gilder's Law states that communications costs will halve and speed will double every 18 months for the foreseeable future. As the number of platforms using a piece of software increases, the fixed costs of producing software can be amortized over a larger number of users. Thus software costs per computer (software power or complexity held constant) have gone down with the growth of installed computer capacity. Moreover, as the market expands, more and more sophisticated software applications can be profitably marketed, thereby increasing options for the user.

---

[3] 'In 1975, it cost one million dollars for the leading computer to process one million instructions. In 1994, a personal computer could process one million instructions for forty-five dollars [2].'

Network economics are also important. The value of a telephone network increases with the square of the number of users.[4] I would suggest that for the World Wide Web, benefits increase even more rapidly with size. Participants on the Web post information as well as search for the information. As the number of hosts connected to the Web increases, then the benefits resulting from each page posted tend to increase to the poster. Moreover, authoring software is amortized over more users, and the costs of each page posted tend to decrease. Consequently, as the number of people on the Web increases, so too does the amount of information each posts.[5] Each surfer benefits not only from the new postings on each new web server, but from the added postings on existing servers. In addition, Metcalfe's Law holds that if you connect any number, n, of computers you get n squared potential value. Thus inherent in Internet technology are increasing returns to increasing network size.

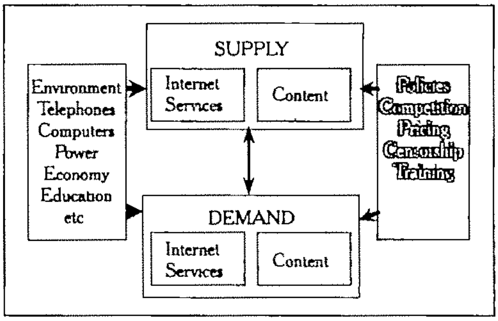
Following Thomas Hughes [7], attention can be directed to 'technological systems' beyond the network. The Internet technological system includes a portfolio of applications (email, listserves, chat rooms, the World Wide Web, e-commerce, Internet telephone, Internet fax, Internet radio, etc.). The evolving physical network makes these applications feasible, while the applications justify the network, each added killer application adds value to the network for its users. Indeed, with the convergence of many information and communications technologies, the technological system of synergistic networks and applications may well extend beyond the Internet to include broadcast media, telephony and computers in general.

These three technological sources of Internet growth are mutually supportive If, for example, reductions in the cost of hardware and software allow Internet services to be offered at lower prices, and the number of subscribers increases correspondingly, then the larger user base will lead to still more benefits per subscriber, and the demand for Internet services will increase. Similarly, if a new killer application comes along, increasing the value of Internet connectivity to the subscriber, more will subscribe, the larger subscriber base will again lead to more benefits per subscriber, and the demand for Internet services will increase still further.

Penetration and utilization

Figure 2 illustrates how connectivity and use of the Internet are addressed. Supply of and demand for Internet services vary with price; an equilibrium price is one where supply equals demand, and that point should establish both the service penetration and the price. In developing countries markets tend to be imperfect and the supply and demand for Internet services might be limited by infrastructure and other variables. Consideration of penetration also involves consideration of

Figure 2. *Penetration and utilization model*



---

[4] Thus the n+1st subscriber benefits from being able to communicate with all n previous subscribers, and each of them benefits from being able to communicate with him; there are 2n units of benefit attached to the n+1st subscription.

[5] Benefits from the n+1st host attached to the net are not 2n units, but rather 2n + an + bn, where a is the amount of benefit each host gets from its added pages posted, and b is the added benefit each host gets from the added pages posted on other hosts.

the quality of connectivity in terms of bandwidth, service quality, and reliability.

There is an old saying, 'You can lead a horse to water, but you can't make it drink.' Similarly, you can connect a man to the Internet, but you cannot make him email or surf. It is not just the penetration of the Internet that seems most likely to determine its impact, but also its use or non-use. While those without access to the Internet can be assumed not to use it, the utilization of Internet capacity by those with access cannot be assumed.

The term 'content' normally refers to the content posted on the World Wide Web, although by extension it might be used to refer to the content of email traffic and other communications via the Internet. Currently most web content is available to the surfer without cost. Nevertheless it seems likely that 'market' considerations can be useful in discussing content. Thus it would seem likely that the larger the number of readers likely to browse a website, the more willing would be the owner of the site to post material. As the ability of buyers and sellers to find each other in a market is of concern, so is the ability of those posting material on the Web to reach relevant readers and the ability of surfers to find that which they seek.[6]

The *pattern* of penetration and utilization of the Internet is also of interest. What are the relative importance of academia, the business sector, government and civil society? What portion of users are foreign, and what nationals? Does utilization of existing connections differ among these groups?

The framework suggests that the utilization of the Internet by those that are connected is influenced by cultural, social and economic factors, and can be influenced by policies. Thus, in Africa, Internet and telephone charges are often per unit time connected and are high, so that cost is likely to limit use. Use is higher in universities and NGOs than in government, and this may be related to organizational climate. Low levels of education and widespread lack of experience with information based decision making also seem likely to lead to low levels of Internet utilization.

Institutional impacts

Many studies deal with institutional impacts of the Internet, several emphasizing impacts in developing countries. Discussions of measurement of impacts on organizations include impacts on organizations in general [9], and on specific kinds of organizations such as (governmental) organizations involved in agriculture [10], civil society organizations [11], relatively large developing country enterprises [12], small and medium enterprises [13], and micro and small enterprises [14]. Similarly, discussions have focused on the impacts at the project level [15]. The staff of the Acacia Project is conducting a study of the impacts of telecenters in Africa [16]. A recent National Academy of Sciences report included discussion of the impacts at the level of the household and community [17]. More general discussions of social and economic impacts exist [18]. More general frameworks for studying the impact of information [19, 20] can be modified to focus on impacts of the Internet.

In developing countries, impacts of the *use* of the Internet should be greater than the direct employment and business impacts of the provision of Internet services. The impacts of non-use of the Internet (notably in Africa) may be even more significant.

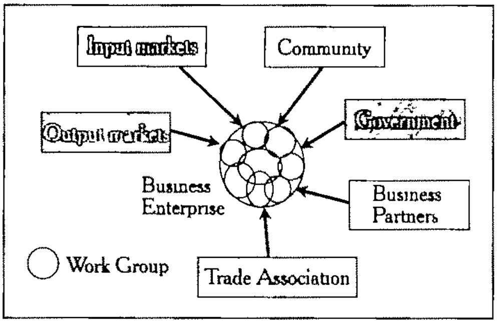
Institutions of concern include formal organizations, such as businesses, schools, hospitals and health centers, non-governmental organizations, and associations. There are of course many other institutions (which are not formal organizations) which can be affected by the Internet. These would include markets, communities, and families.

Figure 3 exemplifies how the conceptual framework deals with institutional impacts using a business enterprise as an example. The enterprise is seen as a system composed of a number of component parts. The diagram suggests a relatively small organization, composed of a number of interrelated work groups (e.g. departments). A reductionist

---

[6] Indeed, search engines may not describe a considerable portion of the World Wide Web according to Lawrence and Giles [8].

Figure 3. *Business enterprise and some relevant institutions*



approach to studying the direct impact of the use of the Internet in such an organization would look at the use of the Internet in the various work groups. Typically, the impacts sought would be relevant to any or all of a number of parameters: efficiency of resource use, effects on organizational functions, or accomplishment of work group objectives. The direct impact on the overall organization might be inferred from the aggregate of work group impacts, and from the structure and processes of the organization. A further impact might be sought in the ways in which the structure and process of the organizations themselves adapt or are modified as a result of the introduction of the Internet.

Similarly, the measure of the direct impact of the Internet on a market depends on its impact on the entities that trade in the market, on market structure and processes, and on how the structure and processes are modified in response to the Internet. Impacts on many other institutions can be considered including the family, community, civil society, political, governance, industries, and professional associations and networks.

With time, institutions will self-organize to adapt to the Internet. Such self-organization may be both:

- teleological or planned, as when executives of a group of enterprises develops a network strategy for those enterprises to collaborate via the Internet, or when entrepreneurs create new enterprises to accomplish social tasks in new ways made possible by the Internet; and
- teleonomic or unplanned, as when feedback in the form of increased profits and investment are used to increase the size of an enterprise successfully utilizing the Internet, or when economic selection by markets results in the disappearance of enterprises which no longer compete effectively in an Internet environment.

Indeed, the very idea of the institution may change; for example, the business enterprise may eventually be seen less as a free standing entity defined by relatively closed boundaries, and more as an entity with open-boundaries which participates in various networks of producers and consumers [21].

Figure 3 also suggests that the behavior of the enterprise is influenced by other institutions. Thus the owners', managers' and employees' reaction to the Internet may be strongly affected by:

- the signals that they detect in input and output markets;
- support from community institutions (such as chambers of commerce);
- support from associations (such as trade associations);
- government policies and services;
- by the actions of partner enterprises (suppliers, customers).

Indeed, an enterprise might be very strongly affected by the Internet through these other institutions even though it itself did not connect or utilize the technology. If, for example, its competitors use the Internet to obtain advantages in input and output markets and to increase productivity, a firm not using the Internet may find itself at a severe competitive disadvantage, and indeed be forced out of business.

Institutions are seen as permeable. Each person, of course, participates in many institutions, and the experience with the Internet in one institutional role can affect other institutional roles. Thus the worker who learns to use the Internet at the office may subsequently use the Internet to participate in activities of his professional society, to communicate with members of his family, and to participate in chat rooms discussing politics. Increased labor mobility engendered by intra-family use of the Internet could affect enterprises by changing labor markets.

As described above, understanding of the overall impact of the Internet will develop from understanding many individual studies of specific impacts, and how they fit together. However, for such an approach to function well, it is important that studies within institutions include descriptions of the relevant behavior of other institutions. Thus, studies of the impacts of the Internet on firms should include considerations of the markets in which those firms acquired resources (primary and intermediate goods, labor, financing) and sold their goods and services, relevant government policies, and the action or lack thereof of chambers of commerce and trade associations.

Specifically, the conceptual framework recognizes that the response of an institution to the Internet is conditioned by policies and by cultural, social and economic factors. Thus, the ways schools are affected by their access to and use of the Internet might be affected by policies of the Ministry of Education, as well as by factors such as the openness of the society, the general level of education, the participation of families in school activities, and the willingness of the private sector to encourage schools to utilize the Internet.

Sectoral analysis

How will the Internet affect:

- economic growth;
- the distribution of wealth and income;
- disease, disability, discomfort and death in the population;
- the ability of people to participate in their own governance;
- food availability and nutrition;
- the knowledge people command and utilize;
- etc.

Often, perhaps usually, individual policy makers are interested not in the entire spectrum of such issues, but in a single one.

The proposed conceptual framework allows for the selection of a group of institutions for each such issue. Thus, if one is interested in the impact of the Internet on public health, one would focus on institutions in the health sector: a ministry of health, health service systems managed by non-governmental organizations, private organizations offering health services on a fee-for-service basis, the markets for pharmaceuticals, medical supplies, health care finance, and labor in the medical fields, and associations of health professionals.

It should be noted that this separation of a system or sector for analysis from its background of the society in general is an analytic convenience. Continuing the example of the health sector, it is well known that the health of populations is not only, or perhaps even primarily, a function of the health services industry and its associated institutions. An affluent, well-fed, well-educated population living in a hygienic environment will generally be healthy, while a poor, undernourished, poorly educated population living in filth will generally be unhealthy in spite of the efforts of its health workers. Thus the impact of the Internet on health may be studied in terms of its effects on the performance of institutions in the health sector, but the effects of the Internet on economic growth, food production and distribution, education, and the environment may ultimately be more significant for health.

Impacts on people

Institutional impacts of the Internet are a useful topic for study because of the considerable importance of teleonomic processes in institutions. Institutions may exhibit orderly behavior that is not planned, but they should not be reified. The impact of the Internet on an institution is not important *per se*. Impacts on people are.

The conceptual framework identifies three classes of impacts of the Internet on people:

- impact stemming from direct (consumer) uses;
- indirect, from their participation in institutions;
- indirect, from changes in the performance of institutions of which they are clients.

The direct impact of the Internet on the individual user seems to be the most commonly discussed, probably because it is the simplest to conceptualize and to measure. It seems unlikely to be the most important. Few people spend a large part of their income on the Internet, and one must assume that the marginal net benefits of such use are comparable to the cost of Internet connectivity.

The indirect impact of the Internet on a participant in an institution may be quite significant. Consider, for example, the employee who loses his job because his company is unable to compete with others that use the Internet more effectively and goes out of business. Or consider the stockholders in those companies that used the Internet effectively to increase productivity and profits, and thus enhanced their stock values, or the bonuses likely to have gone to managers of such companies.

Indeed, the third class of benefits may be still greater. If, for example, power users of the Internet in biomedical research are able to develop vaccines by using their computer power more rapidly, or indeed able to develop vaccines that would not have been possible without such computer support, the benefits to the community may be immense.[7]

In the previous section of this paper, it was suggested that non-use of the Internet might result in serious negative consequences. The examples above also illustrate the possibility that recipients of important benefits from the Internet may not themselves be Internet users, or indeed even be aware of the Internet use from which they benefited. Thus, the beneficiaries of new vaccines developed through power applications of supercomputers and wide bandwidth communications media are quite unlikely to be aware of the role of information and communications technology in vaccine development, much less to have been involved themselves in such uses.

Therefore, a complex pattern of distribution of Internet services is probably required to maximize benefits. Such a pattern that may give priority to Internet II connectivity for power users in critical applications areas, rather than to wide distribution of basic connectivity to consumers. Indeed, it might even be that greater benefits would come to Africans from high level connectivity for vaccine researchers in Europe and the US than from broad penetration of Internet services to households in Africa itself.

Technological determinism versus induced innovation; the Matthew principle versus deterministic policies

The conceptual framework (Figure 1, above) identifies technological change as one of the factors responsible for the rapid increase in the penetration and use of the Internet. However, the conceptual framework also suggests that the penetration and utilization of the Internet and the institutional impacts of that penetration and use are dependent on environmental, social, and economic conditions, and on the policies in place. This situation leads naturally to consideration of the Internet as an 'induced innovation' [22]. The Internet is certainly best viewed as having been invented in developed countries, and the technology transferred to developing countries (from the point of view of developing countries it is *not* an 'induced *invention*'). However, the argument above suggests that the *innovation*[8] is to a significant degree induced by factors internal to developing countries. Essentially this framework rejects a technological determinist position that the Internet is an 'autonomous' technology, rather placing on developing country policy makers the responsibility for creating an environment that induces a pattern of innovation that benefits the populations they serve.

Observed impacts should also be attributed to:

- policies influencing Internet availability, use and non-use and the ways institutions allocate costs and benefits of the Internet, versus

---

[7] Worldwide, nearly eight million people die each year from AIDS, malaria, and TB. In each case vaccines appear possible in the future, and control of the disease appears impossible without a vaccine. Thus, an average of a year saved in developing vaccines for these three diseases alone would be expected to save eight million lives!

[8] An 'innovation' involves the penetration and use of an 'invention' in the society.

- environmental factors that influence institutions to benefit or to fail to benefit from the Internet capacity provided and its use.

Clearly, richer, more educated groups and countries have more access to the Internet and are likely to benefit more from it [23]. Thus, the Matthew principle applies to the Internet.[9] Yet it is clear that good policies and the development of appropriate institutions in developing countries can promote more rapid penetration of the Internet, its more appropriate utilization, and, thus, more beneficial and equitable impacts. If the Matthew principle governs and the 'digital divide' widens, how much of the gap should be considered as a further repercussion of the poverty of developing nations, and how much should be attributed to the failure to develop adequate policies and institutions? While the question is essentially unanswerable, it should be noted that more generally development success has been strongly linked to good policies and good institutions [24].

The Internet in the developing world

Five-sixths of the people of the world live in low and middle income countries; the average per capita GDP for two billion people who live in low income countries was $350 in 1997, and for the nearly three billion who lived in middle income countries it was $1,890. Infant mortality in middle income countries was six times, and in low income countries sixteen times that of high income countries. Low and middle income countries benefited comparatively little and late from technological advances of the past. Even while the Green Revolution revolutionized agriculture in Asia, Africa saw decreases in per capita agricultural production. Developing countries have the world's greatest needs for the advances promised from the Internet, as well as the world's least prepared physical, institutional and human infrastructure to realize these advantages.

It was estimated in May 1999, that there were 171.25 million people connected to the Internet worldwide [25]. Assuming a global population of six billion, this means that more than 97% of the people in the world are *not* connected to the Internet. Most of the wired are in developed countries. Mike Jensen estimates that in Africa (excluding South Africa) there are 4,100 people per Internet user [26]; Africa meets the Ivory Snow[10] test: it is more than 99.9% uncontaminated by the Internet. Thus, the impacts of non-use of the Internet may well predominate in developing countries and in Africa in particular.

Penetration differs radically from country to country. Figure 4 is based on data prepared by Mike Jensen for Internet connectivity in Sub-Saharan Africa [26]. It relates Internet connectivity to per capita GDP.[11] The upper regression line is for countries which are the focus of the Leland Initiative, and the lower for those which are not. The Leland Initiative [27], which seeks to promote the penetration of the Internet in Africa, focuses on countries which appear promising for its purposes, and then engages with their governments in a policy dialog promoting privatization and competition among Internet Service Providers (ISPs), cost-based pricing of telecommunications services for the Internet, and freedom of speech in terms of Internet content. The Leland Initiative has provided subsidies for Internet hardware to countries with which it has signed memoranda of understanding about Internet policies, and has conducted training of (potential) Internet users. The figure suggests not only that Internet penetration increases with increased per capita GDP, but that policies advocated by the Leland Initiative are conducive to increased penetration.

Rob Schware summarized data from a 1997-98 survey by the IDRC's Capacity

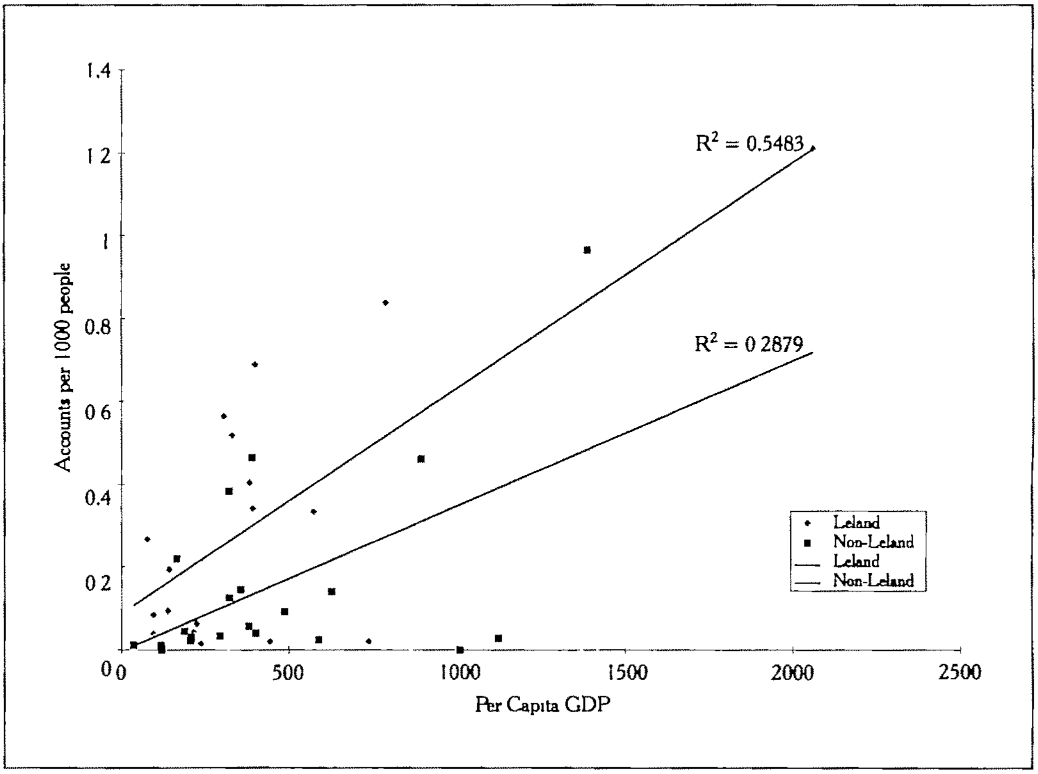
---

[9] 'For whosoever hath, to him shall be given, and he shall have more abundance: but whosoever hath not, from him shall be taken away even that he hath.' (Matthew xiii.7)

[10] A US soap which used to advertise as being 99 and 44/100th per cent pure.

[11] Data are restricted to those countries with per capita GDP in 1996 less than $3,000 and fewer than two Internet accounts per 1000 population.

Figure 4 *Internet accounts vs per capita GDP*



Building for Electronic Communication in Africa (CABECA) project as follows:

> The survey found that there was as yet little government use of the Internet: only one per cent of users in Ethiopia, and only six per cent in Zambia, were government employees. The highest number of users belonged to non-governmental organizations (NGOs), private companies, and universities. The ratio of nationals to non-nationals varied between countries: 44 per cent of users surveyed in Zambia were nationals as compared to 90 per cent in Ghana.
>
> The survey findings suggest that the average level of email use is one per day, sending and receiving. Email is used for general correspondence and document exchange, technical advice, managing projects, arranging meetings, and exchanging research ideas, although its use is still limited for finding and accessing formal information resources. ... The WWW is still a relatively underutilized resource, although 40 per cent of Zambian users questioned had conducted literature searches on the Web [28].

Thus use of the Internet in Africa by those connected is still quite low. For comparison, the Neilsen/Netratings panel reports an average of two hours, forty-one minutes per week surfing the Web in the US [29].

A second comparison can be made based on the Computer Command Index (see the questionnaire presented in the Appendix). The questionnaire was pretested with volunteers belonging to listserves on topics of Internet use and impact, and the sample can be considered to be very active and quite sophisticated users from developed countries. Twenty-six responses are summarized in the table below for all items (except the time using computers, for which there were 16 responses). The scoring method gives five points for a Pentium or equivalent computer as the primary computer, and six points for a Pentium II or equivalent, indicating that these respondents were all using Pentium or more powerful computers.

Item	**Average Response**
Years using computers	14.41
Primary computer	5.15
Other computers in use	1 77
Stand-alone applications in use	4.58
Hours/day stand-alone applications	4.81
Internet applications in use	3 65
Hours/day Internet applications	3.15
Total applications	8.23
Total computer hours/day	7.96

Thus, these people were using computers nearly eight hours per day on average, including using the Internet more than three hours per day.[12] Moreover, they were using not only email and the World Wide Web, but (on average) more than one other Internet application. It seems highly unlikely that there are many such users in Africa.

In developed countries, some direct indicators of the impact of the Internet involve the growth of investment in and revenues from Internet infrastructure, the GDP of the industry manufacturing such infrastructure, and Internet related employment (network managers, web page designers, etc.). Such analyses are less relevant in developing countries, which will generally not manufacture hardware for the Internet, and which have far sparser penetration. Still, even Sub-Saharan Africa has more than one million people connected to the Internet; if we assume that annual cost is some $800 per connection,[13] the Internet budget for the continent would be $800 million. Employment generated by this $800 million per year industry would be significant.

The conceptual framework stresses that the impact of the Internet is understood in terms of institutions, and international institutions must also be considered. Easier said than done! For example, the researcher seeking to understand the impact of the Internet on small and medium size enterprises in West Africa might need to understand how the Internet is affecting world markets in ground nuts, crafts, and tourism services – not an easy task.

Thus, as is well known, financial markets have become increasingly international over the past several decades. It is estimated that business-to-business e-commerce will increase from $43 billion in 1998 to $1.3 trillion in 2003 (while business-to-consumer e-commerce increases from $8 billion to $108 billion in the same period) [31]. Markets for intermediate and final goods will be increasingly international (and mediated via the Internet). Labor markets are international, with many developing countries experiencing diasporas of their (educated) workers, and many expatriates from developed countries working in developing countries. Many businesses and NGOs are multinational, and governments are involved with international agencies. Scientific and technological networks are often global.

The knowledge and information gradient between developed and developing countries is steepening. As African countries are increasingly competing with more wired countries in Asia and Latin America, and facing increasing knowledge gaps as compared with developed countries, the impacts of low levels of use or non-use of the Internet appear likely to be severe. Sophisticated Internet users seek to develop markets in Africa, Asia and Latin America for developed country products, as developing countries seek to develop markets in Europe and North America for their products. While increased trade offers great potential for developing countries, there is at least a risk that the superior Internet sophistication of Northern business will engender worsened terms or trade or increased trade deficits for the South.

---

[12] Note that respondents had little difficulty in separating stand-alone use of their personal computers from Internet based uses, and that they were using more stand-alone applications and using them more than Internet applications. These data may suggest that the Internet still has less impact than the PC, even on sophisticated users.

[13] The annual cost for South Africa was estimated at $793 and other countries would be higher [30].

Many other institutions are also (but less obviously) international. Families and communities use the Internet to communicate internationally.[14] Many ethnic groups are multinational, not only from migration, but also as a result of national boundaries that cross-cut the regions in which the ethnic communities live, and the traditional institutions within these ethnic groups are similarly likely to transcend national boundaries. The underground economy is increasingly international and, unfortunately, criminal cartels engage in a global drug trade Terrorists have an opportunity to use Internet technology to achieve secure domestic and international communications of a kind available only to governments in the past.

In Africa, governments have been weak and overburdened by portfolios of government services that exceed their available resources. They will be challenged to utilize the Internet to improve services and productivity, and to involve citizens more in governmental decision making Surely those favoring antidemocratic positions – from technocracy, to demagoguery, to dictatorship – will also see potentials to advance their interests via the Internet. The ability of African governments to assure peace and security may be further eroded as the underground becomes increasingly international and wired.

Some of the most important benefits that will come to Africans from the Internet will be from power applications of which few Africans will be even aware. For example, the Famine Early Warning System (FEWS) [33] applies considerable computer power to remote sensing data, as well as a variety of other social and economic data to identify areas in which food shortages are likely to occur. By providing advance warning it allows governments to take cost-effective steps to avert famine. Note that in this case a relatively small number of power users of information and communications technology are able to develop information, and a relatively modest use of the Internet in Africa allows this information to be rapidly disseminated to key decision makers. A comparable application exists using information and communications technology to predict epidemics [34, 35]. Computer analysis of weather conditions based on remote sensing data allows identification of conditions likely to foster epidemics of rift valley fever and other infectious diseases of men and livestock. The Internet requirements needed to distribute timely warnings to public health and veterinary health officials are modest.

Donor agencies

Foreign assistance is an important factor in the transfer of Internet technology to the developing world. It seems likely that most foreign assistance agencies promote the use of the Internet within their own programs and among their partners; the technology is pervasive in the developed countries in which they are based, and contribute to the quality and efficiency of their programs. Some donor agencies have undertaken specific programs to promote the development of the Internet in developing countries, and especially in Africa [36].

There are some programs that have sought specifically to develop case studies and similar materials on the impacts of the Internet [37]. Importantly, the Acacia Initiative [38] (a $60 million project of the Canadian IDRC), the Leland Initiative (described above) and the *info*Dev [39] program (supported by a consortium of donors and managed by the World Bank) have agreed to co-operate in their evaluation programs to develop a large body of information on the lessons learned from donor programs, and jointly sponsored a meeting to develop common understandings of methods in the autumn of 1999.

---

[14] For example, a wedding in Virginia was watched on 'TV on the Web' by people in Calcutta, Bombay and New Delhi [32].

Closing comments

While penetration and use of the Internet are much less in developing countries than in developed, the impacts of the use of the Internet are likely to be important in developing countries. The impacts of non-use of the Internet in an increasingly wired world are also likely to be important.

It seems clear that many of the impacts will be quite positive. The Internet holds great promise for developing countries in communicable disease control It is an obvious medium for study of environmental data and communications relevant to environmental policies. It offers great potential for the dissemination of policy relevant information to governments and the general public. It provides an obvious medium for the dissemination of educational materials and content, and for co-ordination of educational administration. It could be used to improve food supplies, notably in improving the efficiency of food markets and in disaster early-warning systems. And it could enable developing country enterprises to compete more effectively for markets abroad. Many of these potential benefits are likely to be shared by entire populations, and indeed to benefit the poor and needy even more than the affluent.

However, there are also threats to developing countries. It is not hard to imagine Internet scenarios in which brain drain is increased, financial crises occur more often and in more damaging ways, more wired countries use their information advantage to unfairly exploit the natural resources of developing countries, undesired transfers occur from cultures of countries with more content on the Web to cultures within developing countries, crime is conducted via the Web, insurgents use the Internet effectively in their efforts to destabilize democratic governments, developed country businesses use e-commerce to sell to developing country consumers (thereby increasing capital flight and diverting scarce capital to elite consumer goods rather than growth enhancing investment), or those already benefiting from relative superiority of wealth, power or education are more successful in appropriating the technology than their less advantaged compatriots and use it in ways that increase inequities in their societies.

The case has been made that the penetration of the Internet will depend on policies and environmental factors in the developing countries, that the use of the installed Internet capacity will also depend on policy and environmental factors. It has been stated that the effects of use and non-use will also depend on policy and environmental factors It is suggested that it is important to distinguish between the impacts that are inherent in the nature of the Internet, those which are subject to policy control, and those which vary from country to country as a result of factors beyond the control of Internet policy makers.

If the promise of the Internet is to be realized in developing countries, if the threats resulting from the development and dissemination of the Internet are to be ameliorated or avoided, and if the net effect of the introduction of the Internet in developing countries is to enhance equity and contribute to social and economic development, then policies will have to be good and effective. A program of studies is needed to develop the understanding on which such policies can be based. Fortunately, a consortium of donor agencies is beginning a co-ordinated effort to develop such a program.

References

1. Standage, T. *The Victorian Internet: the remarkable story of the telegraph and the nineteenth century's on-line pioneers* New York: Walker and Company, 1998, 1-2

2. Huseman, R.C. and Goodman, J P. *Realm of the Red Queen: the impact of change on corporate structure, corporate education, and the emergence of knowledge organizations.* 7. http://www.ec2 edu/kworld/knowledge/paper/realm.pdf

3. Quoted in: Adams, R. McC. *Paths of fire. an anthropologists inquiry into western technology.* Princeton University Press, 1996, 6.

4 Menou, M. Impact of the Internet: some conceptual and methodological issues: or how to hit a moving target behind the smoke screen. In: Nicholas, D. and Rowlands I., eds. *The Internet: its impact and evaluation. Proceedings of an international forum held at Cumberland Lodge,*

Windsor Park, 16-18 July 1999. London Alsib, 2000

5 Schware, R. et al. *Economic Internet toolkit for African policy makers.* 1998. http·//www.worldbank.org/infodev/projects/afprelim pdf

6. National Academy of Sciences. *Internet counts: measuring the impacts of the Internet* Washington, DC, 1998 http://www.bsos.umd.edu/cidcm/wilson/xnasrep2.htm

7. Hughes, T.P. *Networks of power. electrification in western society (1880-1930).* Johns Hopkins University Press, 1983.

8. Lawrence, S. and Giles, C L. Accessibility of information on the Web. *Nature*, 400(6740), 1999, 107.

9 King, J.L. and Kraemer, K L. Computer and communication technologies: impacts on the organization of enterprise and the establishment and maintenance of civil society. In: National Academy of Sciences. *Fostering research on the economic and social impacts of information technology: report of a workshop.* 1998. http.//www.nap.edu/readingroom/books/esi/index.html

10. *Assessing the impact of information and communication management on institutional performance. proceedings of a CTA workshop.* Wangeningen, the Netherlands: CTA, 1998.

11. Measuring the impact of the Internet on civil society organizations in Central America: research proposal – project concept *Fundación Acceso*, San José, Costa Rica, March 1, 1999.

12. Daly, J. and Miller, R.R. *Corporations' use of the Internet in developing countries.* International Finance Corporation Working Paper Number 35, Washington DC, 1998. http·//www.ifc.org/DEPTS/OPS/ECON/PUBS/DP35/Dp35.htm

13. Lefebvre, É. and Lefebvre, L.A. *Information and telecommunication technologies. the impact of their adoption on small and medium-sized enterprises.* IDRC, 1996. http://www.idrc.ca/books/focus/807/

14. Barton, C. and Bear, M. *Information and communications technologies: are they the key to viable business development services for micro and small enterprises?* Microenterprise Best Practices Project, March 1999, Appendix II. http.//www.mip.org/pubs/MBP/ict.htm

15. Lanfranco, S *A meta-level conceptual framework for evaluating projects involving information and communication technology (ICT).* Bellanet, http.//www.bellanet.org/partners/ldia/lessons/evalfram.htm

16 Whyte, A. *Telecentre research framework for Acacia.* Ottawa: IDRC, 1998. Also available at. http://www.idrc ca/acacia/stcat.htm

17 *Fostering research on the economic and social impacts of information technology: report of a workshop.* Washington, DC: National Academy of Sciences, 1998. Also available at: http·//www.nap.edu/readingroom/books/esi/index html

18. C.f. Collins, E. *Charting the impacts of information technology* Washington, DC: National Science Foundation, 1997. Also available at: http.//www nsf.gov/sbe/srs/infotech/charting/start.htm and *Conference on Information Technologies and Social Development.* United Nations Research Institute for Social Development (UNRISD), Geneva, 22-23 June 1998.

19. Menou, M., ed. *Measuring the impact of information on development.* Ottawa. IDRC 1993

20. McConnell, P. *Making a difference measuring the impact of information on development: proceedings of a workshop*. Ottawa: IDRC, 1995. Also available at: http://www.idrc.ca/books/focus/783/index.html

21. Survey, business and the Internet. *The Economist*, 26 June – 2 July 1999, 17.

22. Binswanger, H.P , Ruttan, V W. et al. *Induced innovation: technology, institutions and developent.* Johns Hopkins University Press, 1978.

23. Daly, J. Infrastructure for whom. *iMP*, September, 1999.

24 *Assessing AID· what works, what doesn't, and why.* A World Bank Policy Research Report. 1998. http.//www.worldbank.org/research/aid/aidtoc htm

25. NUA Internet Surveys. http://www.nua ie/surveys/how_many_online/index.html

26. Jensen, M. *African Internet status· May 1999.* http.//www3.wn.apc.org/africa/afstat htm

27 The Leland Initiative. http://www.info. usaid.gov/leland

28. Schware, R. et al. *Economic Internet toolkit for African policy makers* 1998, Chapter 3. http://www.worldbank.org/infodev/projects/afprelim pdf

29 The Neilsen//Netratings Reporter, June 24, 1999.

30. Schware, R. et al. *Economic Internet toolkit for African policy makers.* 1998, Appendix 6. http://www.worldbank.org/infodev/projects/afprelim.pdf

31. Survey, business and the Internet. *The Economist*, 26 June – 2 July, 1999, 6.

32. *Washington Post*, Monday 28 June 1999, B3.

33. Famine Early Warning System (FEWS). http://www.info.usaid.gov/regions/afr/fews/http://www.info usaid.gov/regions/afr/fews/

34. See for example: Linthicum, K.J. et al. Climate and satellite indicators to forecast rift valley fever epidemics in Kenya. *Science*, 285, 1999, 397-400.

35. See for example: Epstein, P.R. Climate and health. *Science*, 285, 1999, 347-348.

36. Jensen, M. Summary of international ICT development projects in Africa. http://www3.wn.apc.org/africa/projects.htm PICTA: Partnership for Information and Communication Technologies in Africa. http://www.bellanet.org/partners/picta/ and AISI: African Information Society Initiative. http://www.bellanet.org/partners/aisi/partners.htm

37. AISI Case studies: http://www.bellanet. org/partners/aisi/proj/studies.htm The International Institute for Communications in Development: http://www.iicd.org/stories/

38. Canadian IDRC, Acacia Initiative. http://www.idrc.ca/acacia/

39. World Bank, infoDev Program. http://www.worldbank.org/InfoDev

Appendix
Computer command index questionnaire

The purpose of this questionnaire is to obtain information on how much of a command of personal computer technology you have. It is divided into three sections. Part A looks at the hardware you command. Please choose the computer that you use most as the basis of you answers to questions 1 and 2. Part B looks at your use of personal computers as 'stand-alone' devices. Part C looks at your use of the Internet.

General:

1. In what country do you work? ___________

2. How long have you been using computers? ___ years

3. The following answers apply to the computer you use most which is your:

[ ] Home Computer
[ ] Office Computer

Part A—Hardware

A-1. To what generation of hardware does the computer you most use belong?

[ ] 8086 PC or equivalent. (1 point)
[ ] 80286 PC, 68020 Macintosh or equivalent. (2 points)
[ ] 80386 PC, 68030 Macintosh or equivalent. (3 points)
[ ] 80486 PC, 68040 Macintosh or equivalent. (4 points)
[ ] Pentium PC, Power PC Macintosh or equivalent. (5 points)
[ ] Pentium II PC, G3 PC Macintosh or equivalent. (6 points)

A-2. In addition to the computer described in question A-1, how many other computers do you use regularly? [ ] (one point for each additional computer).

Part B-Stand-alone applications in use (i.e. not requiring the Internet)

B-1. Check for each application you **use** at home or at work:

[ ] Word Processing
[ ] Spreadsheet
[ ] Presentation Manager (e.g. power point)
[ ] Data Base Manager
[ ] Calendar/Scheduler
[ ] Desktop publishing
[ ] Financial analysis package
[ ] Other (games package, project manager, health advisor, etc.)
(one point for each box checked)

B-2. Average number of hours per day you use these applications at home and at work in total:
[ ] Round to the nearest hour. (one point for each hour)

Part C-Networking applications in use (requiring the Internet)

C-1. Internet applications you use at home or at work:

[ ] Do you use email?
[ ] Do you search the World Wide Web?
[ ] Do you have, create or manage web pages?

[ ] Do you use one or more streaming applications (Internet phone, radio, video, etc.)
[ ] Do you use one or more agents?
[ ] others (chat rooms, Internet fax, etc.)
(one point for each box checked)

C-2. Average number of hours per day you use these networking applications at home and at work in total:
[ ] Round to the nearest hour. (one point per hour)

Information skills training in support of a joint electronic library in Glasgow: the GAELS Project approach to library courseware development

Nick Joint and Bob Kemp
Centre for Digital Library Research, University of Strathclyde, Glasgow. G4 0NS.
n.c.joint@strath.ac.uk

Susan Ashworth
Glasgow University Library, Glasgow. G12 8QE.

The GAELS Project is a two-year project funded by the Scottish Higher Education Funding Council (SHEFC) strategic change initiative, which promotes collaborative information services to engineering researchers at Glasgow and Strathclyde Universities. This paper examines the role of user education in this process. We use arguments against the effectiveness of library skills education and evaluative methods learned from human-computer interface design as a means of improving information skills training and as part of a general reflection on user education and library services. Such an approach shows how networked learning materials can be an effective tool for promoting a collaborative library service across the Glasgow Metropolitan Area Network.

Introduction

What is the role of user education in the modern library and how should the teacher librarian develop that role in the immediate future? This article is an attempt to explore this and other related issues, in the light of work performed by a collaborative Scottish university library project based at Glasgow and Strathclyde Universities, the GAELS (Glasgow Allied Electronically with Strathclyde) Project. It is not a detailed project report. It is rather an exploration of the underlying pedagogic issues that contributed to the creation of GAELS and also a reflection on the ideas which the project has generated at its mid-point.

GAELS is a two-year research project, started in July 1998 with funding from the SHEFC strategic change initiative. Its remit includes the promotion of a culture change among engineering academics and researchers at both institutions. This change would see a decreased dependence on the hard copy engineering collections at both sites and increasing electronic use of networked access to full-text resources, possibly, but not necessarily supplied by a local document delivery service from the richer engineering collection at Strathclyde University Library. The project has conducted an initial information audit on the use of information by engineers at both institutions [1] and has designed a set of courseware, which provides engineering research academics with networked information retrieval skills that will make them more independent of their local engineering research collection. One possible outcome of this culture change might be the opportunity to rationalise hard copy holdings at both libraries, which display a fair degree of overlap in core material.

Are library and information skills training misdirected?

The most important preliminary question a library courseware development project must address is whether library user education is a valid activity or not? We found that there is an influential body of literature which argues that user education – be it traditional or courseware-based – is the answer to a problem that should not exist. Rather than teach users to use libraries, make libraries more usable and the need to educate should decrease or even disappear. If there is a need to teach library and information skills, this should be done in the academic teaching departments as part of coursework, not stripped away and performed out of context. This is a particularly potent argument when heard by university managers who rightly wish to minimise costs, such as the cost of user education or courseware development. User education can be done either by departmental lecturers, or not done at all.

Proponents of this case such as Pacey [2] argue for a self-explanatory library in which users 'acquire library skills by using the library: all is explained; everything is transparent'; Savenije [3] points out that library instruction:

> has become indispensable since the use of library services has been made too intricate by the libraries themselves... Computer interfaces are evolving into user-friendly tools, so enhancing the accessibility of information.

Finally, Herrington [4] argues that 'Library systems must be so easy to use and transparent that there is no need for library instruction.' Both Pacey and Savenije also promote the idea of reintegrating information skills acquisition into the context of users' mainstream coursework or research, rather than stripping out user education into the vacuum of free-standing information skills acquisition classes delivered by librarians. Herrington, however, argues simply that the Ohio State University gateway design minimises the need for user education: the result of refocussing instructional librarians on designing user-friendly interfaces, rather than teaching.

If this argument were correct, its impact on a project such as GAELS would be significant. Firstly, our attention would turn to the refinement of both the interface to the services developed by the project, and also the interfaces to established local services. Secondly, we would focus on the issue of support for lecturing staff as they expanded the information skills content of their own teaching. However, there are important arguments against the 'transparent interface' and 'reintegration into the curriculum' school of thought which make the creation of a set of free-standing learning materials, such as the GAELS courseware, still valid.

Counter arguments (1): the myth of total transparency

Firstly, human-computer interaction writers themselves believe that there is no such thing as a transparent user interface. They accept that one must design the best interface possible and then use other support devices (help screens, user education and the like) to compensate for the inadequacies of the interface. Librarians, as outsiders to the interface design process, may overestimate the power of user interface design, unlike the interface designers themselves. The principle of transparency, in Rutkowski's [5] words, means that 'the user is able to apply intellect directly to the task; the tool itself seems to disappear'. However, Shneiderman [6], citing Rutkowski, comments:

> Even though increasing attention is being paid to improving user-interface design, the complexity of online systems grows. There will always be a need for supplemental materials that aid users, in both paper and online form...'

Shneiderman cites as examples of these supplemental materials, brief getting started notes and the online tutorial. Other HCI texts also cite help lines and manuals. The equivalent library materials are our familiar library handouts, library computer-aided learning materials, reference desks and library guides.

What is true of user interfaces is true of the traditional library. There are specific reasons why transparency of use is difficult to achieve in the traditional library. The usability of the physical library is hard to optimise. The hard copy collection may be well laid out at one point in time, but that arrangement becomes dated, and the physical rearrangement, re-signing and guiding of printed collections is costly or impractical. Acceptable compromises are made, and user education or library orientation sessions make up for the shortcomings. By contrast, the electronic library may appear to be easily configured and reconfigured for maximum usability, but most library datasets are externally provided by commercial suppliers, and with this comes the proliferation of interfaces experienced in a typical academic library context (as Shneiderman says 'the complexity of online systems grows...'). Prior to advances on the Z39.50 front, the local library cannot control interfaces that resist being homogenised into a single user interface, transparent or otherwise.

Counter arguments (2): the HE context, academic preference and the FTE economy

The argument for integrating user education back into departmental teaching is undoubtedly powerful, since user education that is not linked to a specific set of tasks easily becomes too abstract to be effective. It leads the anti-user education school to the conclusion that the recent emphasis on information literacy (that is, the teaching not just of information retrieval skills, but also of the goals for which information skills are needed) is an inappropriate response to an unnecessary problem. It is the result of teaching information skills in the library, outside of the meaningful context of coursework and its tasks. This debate over the value of information literacy is well-documented [7].

However, the clear perception among academic engineering staff at Glasgow and Strathclyde was that their information retrieval skills had lagged behind the growth in electronic library services. Just as they were not in a position to move to electronic browsing for research information, neither were they in a position to provide information skills training through their teaching or research supervision. Moreover, library user education is a unique form of teaching activity in British higher education, in that it is one of the few extensive forms of service teaching that is provided without the requirement of FTE payments to the teaching department (the library). Because of this, it takes place without the evaluative filter of teaching quality assessment. This is unlike, for example, engineering mathematics service teaching, delivered by mathematics departments, to engineering departments for FTE income within the structure of TQA. This simple fact of higher education economics does create a financial incentive to departments to accept library and information skills teaching delivered outside the departmental teaching or research context. Fewer departmental resources are involved.

Unfortunately, because such user education is exempt from the teaching quality assessment that departmental teaching receives, it makes the anti-user education argument so easy to promote. If such teaching is never formally evaluated by an outside agency, and if libraries themselves never formally evaluate the quality of their teaching (which is the case according to analyses of library user education evaluation carried out by writers such as Bober [8]), it is hardly surprising that a strong argument can be made against the quality and validity of library-based user education. There is no counter argument being made, no attempt to evaluate and justify this teaching on a wholescale basis.

Decisions on a way forward

Our first decision was that, in practice, freestanding library-delivered user education will continue due to academics' perceptions of their information skills levels and to the nature of user education funding. A consequence of this should be that the evaluative role of departmental teaching quality assessment is replicated in the library, and certainly in a courseware project such as the Computer-aided learning (CAL) strand of GAELS. Secondly, and most importantly, library-led user education should continue because, if done well, it is worthwhile. The difficulty of creating a largely transparent user interface to library services means that compensatory information skills training will be necessary for the foreseeable future. However, interface design methods also demand that evaluation of the quality of your training programmes and library systems constantly takes place.

Bearing these considerations in mind, our challenge was to design a set of learning materials that responded to the following requirements:

- The courseware should be related to the task-content of departmental teaching, but still delivered as a course-independent package.
- It should use the goal of interface transparency as a requirement for the interface to the learning materials, regardless of the lack of transparency of the information services that the package described.
- The package should be created on the basis of evaluative feedback from users, in order to replicate in some way the role of departmental teaching quality assessment procedures and to acknowledge the requirement for evaluation of interface quality as demanded by human-computer interface design methodology

Task-orientation in library courseware

As we have seen, much of the ammunition for the anti-user education school came from ideas that had parallels in user interface design. A fuller consideration of these ideas showed that user education is in fact an inevitable part of interface design. The next question therefore is, can user interface design principles also provide ways of dealing with the valid argument that library user education out of context is inevitably less effective?

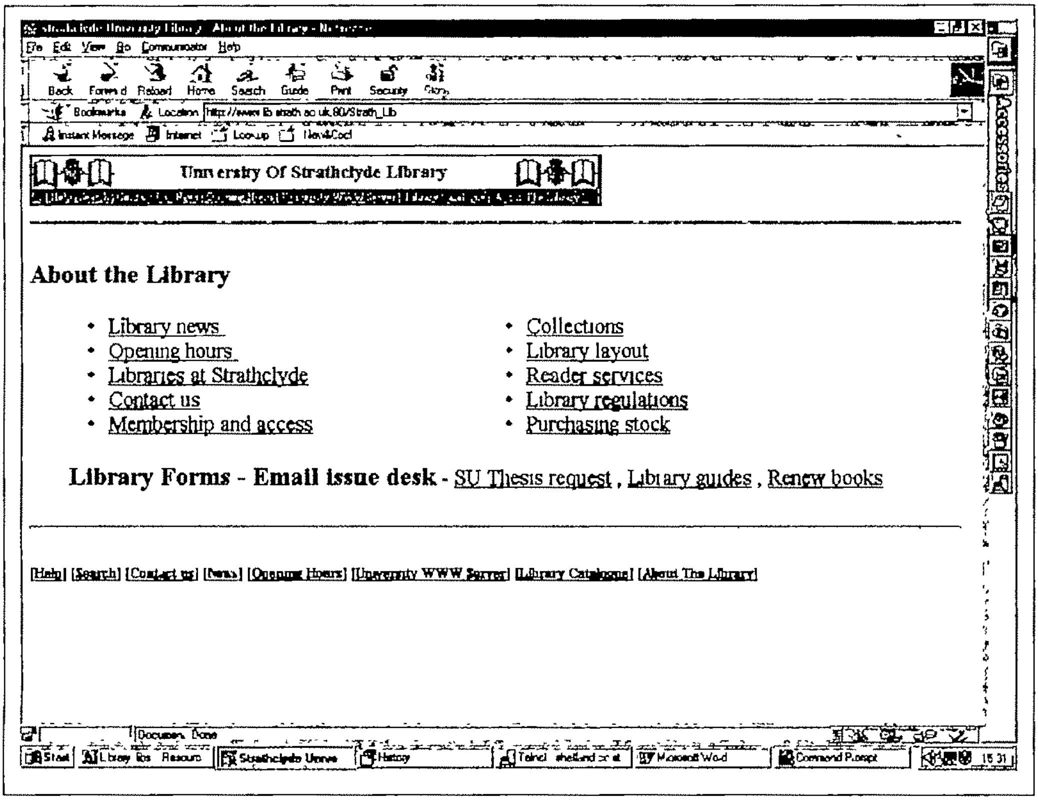
One important principle for making user interfaces as effective as possible is the idea of task orientation, that is the aim of making interfaces task-specific, rather than type- or process-specific. It seems that this principle can give library courseware developers a way of creating a degree of task relevance without embedding learning materials directly into coursework, however desirable that might ultimately be.

Although the literature of task analysis has become complex and rich in the last two decades, for our purposes we could use these methodologies selectively in support of our courseware development. For example, in Lewis and Rieman's [9] description of the interface development process they emphasise the selection and analysis of real, complete, representative tasks for task-centred design as opposed to abstract, partial elements. They also argue that the interface designer should differentiate tasks from processes, that is 'define what the user needs to do, not how it will be done'. What does this mean when applied to library skills training, or library interfaces in general?

If one were to take a WWW-based set of library information resources, the interface to these resources could typically be an opening home page with a choice of hyperlinked paths through these resources. The Strathclyde University Library 'About the Library' WWW pages [10] (see Figure 1) adopt this format They give paths into the data such as 'Opening hours', 'Libraries at Strathclyde', 'Membership and access' and 'Collections of the University Library'. Although such a WWW resource is useful, its interface is not at all task-centred in Lewis and Rieman's terms and lacks usability. In their terms, these hyperlinks are 'abstract, partial elements' in the activity of using a library, aspects of processes familiar to librarians; or they are comprehensive descriptions of processes, in which are submerged the tasks users wish to perform.

Simply renaming these elements would help. For example, the 'Membership and access' link could be re-titled 'Joining the library'. At least then a library user would be able to identify an activity with which they were familiar. But the activity of 'Joining the library' is only a small element in the complete and representative task of (for example) 'finding literature on your topic'. Similarly, the link entitled 'Purchasing stock' is an activity of marginal interest to most library users, and in this instance it leads to a detailed description of the local collection development process. The task of 'finding literature on your topic' may involve some knowledge of this material (for example, it is important to know if a library's collection development policy means that it is likely to collect information on your topic). But it is hard to pick out such information in the overall description of a library process.

Figure 1. *Types, elements and processes· a WWW page from Strathclyde University Library*


File Edit View Go Communicator Help
Back Forward Reload Home Search Guide Print Security Stop
Bookmarks Location http://www.lib.strath.ac.uk:80/Strath_Lib
Instant Message Internet Lookup New&Cool
Document: Done
Start
15:31

University Of Strathclyde Library

About the Library

- Library news
- Opening hours
- Libraries at Strathclyde
- Contact us
- Membership and access
- Collections
- Library layout
- Reader services
- Library regulations
- Purchasing stock

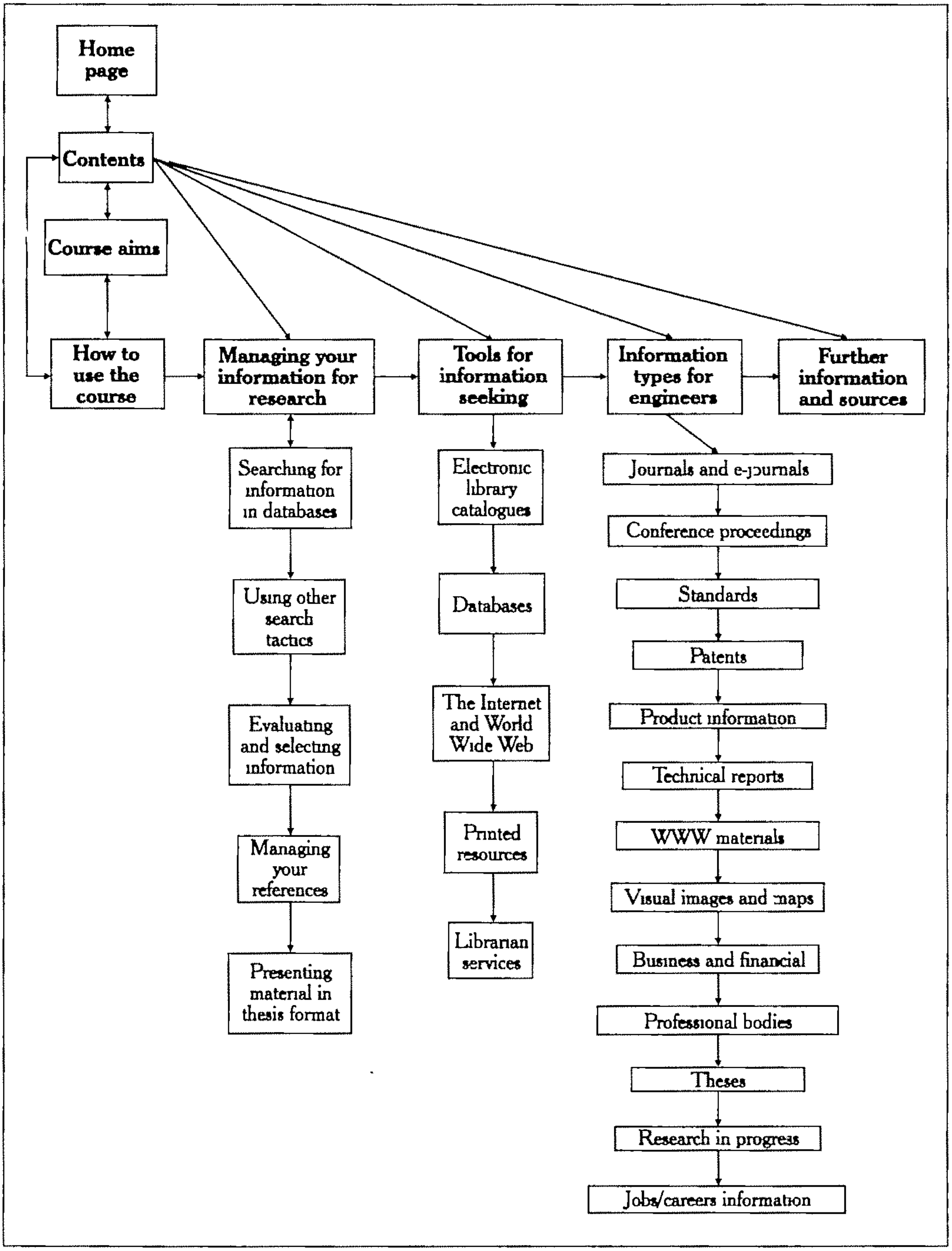
Library Forms - **Email issue desk** - SU Thesis request , Library guides , Renew books

[Help] [Search] [Contact us] [News] [Opening Hours] [University WWW Server] [Library Catalogue] [About The Library]

Such a brief assessment of the task-centredness of a local library interface does a number of things. Firstly, it illustrates the power of the anti-user education argument. In this case, a lack of task-orientation occludes the transparency of a local library interface. A user should be able easily to identify the tasks for which an interface is designed, rather than be presented with a variety of smaller elements of whole tasks or comprehensive overviews of library processes that mask task outlines. Secondly, if this approach does indeed help improve the usability of an information resource interface, it also gives an indication of how to present learning materials in an effective way, with a usable interface.

Developing these insights further, there seemed to be a variety of concrete ways in which task orientation could inform the design of the package. The starting points from which users entered the GAELS courseware could be task-specific. This would mean arranging a gateway of access points to the courseware, each of which is a real, complete, representative task. For example, the Pathfinders Route through the EDUCATE (End-user Courses in Information Access through Communication Technology) Project's *Into Info* [11] CAL package gives the user a variety of named tasks via which to approach the learning materials, such as 'Starting to use the library', 'Searching for facts', and 'Starting research'. Then, at the most specific level of use, the course unit, any one page could start by identifying to the user the task or tasks to be mastered once they had completed the unit. Thirdly, and most obviously, the unit exercises, together with the feedback on the results of the exercises, should involve the execution of real, representative tasks. Lastly, analysis of user needs,

Figure 2. *GAELS courseware structure diagram*



evaluation and user feedback would indicate whether the content of the package was truly representative of the tasks needing to be performed by its target market, engineering researchers.

Librarians' preconceptions about information retrieval tasks

One important consideration for the project was stakeholder commitment. We wanted our project courseware approach to be acceptable to teacher librarians, so that they would use the materials in their outreach work with departments. It was essential to avoid dangers, seen previously in other computer-aided learning projects, where specialist staff used sophisticated skills to craft a product that teaching librarians did not subsequently use. Above all, we wanted to create a sense of ownership of the materials by such teaching subject librarians.

Two problems emerged as a consequence of this requirement. Firstly, at the level of interface development, there was stakeholder resistance to a task-oriented approach in the presentation of user education materials. Secondly, the librarians who were key institutional stakeholders had clear ideas of what information retrieval techniques should be taught, but this syllabus seemed likely to be problematic when evaluated against likely patterns of actual users' task execution.

As a compromise, task orientation at the page level was retained, but a larger structural architecture that mimicked the sequence of the searching process was used. Thus, moving in a linear sequence through the GAELS [12] courseware package, the student first learns to search in Module 1 (e.g. learning Boolean skills, search field qualifiers and so forth). In the next Module they then find out where to go to implement their searching skills (that is, on search tools such as electronic library catalogues and bibliographic databases). Then in Module 3 they find out about the types of information listed by these search tools (theses, patents, grey literature and other types). This structure enacts the sequence of the search process, in that you start by learning how to search, then discover more about where to search and then learn about what you are searching for (see Figure 2).[1]

Evaluation

How did stakeholder preferences for a more process-oriented approach affect learning outcomes as shown through evaluation? We used a number of evaluation measures, including pre- and post-task questionnaires measuring user confidence, usability and usefulness questionnaires, expert review, user observation and subject librarian questionnaires [13]. The use of the GAELS Project courseware to date shows that the compromises we decided to make to a strict task-oriented approach did not lead to unsatisfactory learning outcomes. We noted the following features thrown up by our evaluative work:

- Assessment of learning outcomes is positive. This set of learning materials seems to fulfil the original remit of the project; however, this did not tell the whole picture.
- Evaluation may tell you how well you have taught someone to use your services, but still does not show you how usable those services are. The anti-user education school would say that evaluation in this sense misses the point. It is more important to evaluate the usability of the services taught than the usability and effectiveness of the learning materials that facilitate the use of the services. This charge still had not been laid to rest. Indeed, the interest of the project in interface transparency issues had illuminated exactly what writers such as Herrington had described, questionable local interface design.
- Although evaluation of the effectiveness of the learning materials did not show that the larger process-oriented design structure impeded the usability of the learning materials, it did – as we had anticipated – reflect back on the validity of the syllabus of material taught. Feedback from interviews with research students who completed the course led us to question the correctness

---

[1] This is only an overview of the project at its midpoint. As stated above, a fuller project report will be made available.

of the syllabus of skills, or task-content that we were teaching.

This third point needs to be expanded For example, in the course of collecting evaluative feedback, one research student who had completed the courseware told us the following:

> I've enjoyed learning the searching skills taught in this package and can definitely search better, but I will still get my information the way I did before in preference. I prefer taking a few key references used by my research group and reading the references used at the end of each paper as my bibliography – searching on bibliographic databases gives too many hits and my own way of browsing through a few references means I don't have to work through long lists of less relevant hits.

This research student was clearly using different information retrieval techniques to gather research information from those we were promoting He had enjoyed using our courseware, but felt that the skills he had learned and improved on in the process of using the courseware (broadly definable as information searching skills) were not as relevant to the tasks he was engaged in as skills that he had evolved on his own (broadly definable as information browsing skills). He understood his information retrieval task as something requiring a different strategy from the ones we were giving him. The idea of precision and recall was accepted, because he realised he needed to raise recall from an overly precise core set of hits. But he trusted his own 'reference chaining' strategy more than a search technique such as ours.

Feedback such as this showed that we were successfully teaching the syllabus of information retrieval search skills we had chosen to teach, but that we needed to evaluate not only how well we taught our syllabus, but also whether the syllabus itself was properly composed. This type of evaluative feedback reinforced our suspicion that our syllabus was rather conservative, in order to create acceptability amongst librarian teachers and get the learning materials accepted in everyday teaching practice.

What did the literature tell us about this problem? Our background reading to the project included writers such as Bates [14] and Marchionini [15] who have described alternative information retrieval techniques as demonstrated in actual user behaviour However, we were interested to read articles such as those by Dyckman [16], who described a process-oriented approach to teaching end-users search skills on the Dialog online host service. This is a valuable and insightful article. It shows a practitioner librarian seeking evidence of actual information retrieval behaviour by users, and re-evaluating their assumptions in the face of this. The article includes a description of end-users generating and accepting high recall, low precision search results:

> Another technique used by many of NYU's [New York University] DIALOG searchers was shocking at first to librarians. Analysis of DIALOG Classmate invoices showed many searches, which retrieved numerous results. It seems that many want great recall and are little concerned with precision. The idea of the 'perfect 20 item search', developed in response to systems which penalize large retrievals (i.e. charge by item and time expended), is incompatible with academic end-user searchers' perceptions of their needs.

Dyckman concludes that searchers need comprehensive recall, not precise recall. Our initial analysis of our user feedback leads us to suggest that perhaps they do want precise recall, but, as in the NYU case study, simply want to browse through search results in order to create a precise set. Browsing large sets enabled them to control personally the selection or de-selection of relevant hits. Had they used search techniques, the reduction of hits would have been done by an impersonal machine process and there would have been a loss of this sense of personal control. This sense of personal control justified the time invested in an extended browse, rather than in a quicker machine-executed search.

Thus, if our research students use browsing techniques to increase recall, why should end-users in other studies not browse effectively to reduce recall? It would have been interesting to have asked end-user searchers in the NYU case study what they did do with their overly long sets of results. Did they in fact browse them to extract a core of relevant references? The search logs themselves would not have shown this process of browsing to increase precision, the author simply assumed that the search log contained a complete task analysis. Because Dyckman did not perform such interviews, the assumption that comprehensive recall alone was desired may have been correct. But evaluative user feedback would have clarified the matter further, thus underlining its important role.

Both the NYU case study search log analysis and the GAELS Project's investigation of actual user information retrieval activity show the same pattern. Users are taught search skills and then fail to search 'properly'. The teacher may then assume that the syllabus of information retrieval skills has been taught badly. However, we suggest that this is probably a misunderstanding of the evaluative feedback. Rather, it is likely that the syllabus has been taught well, but that the syllabus is not correct, or not wide enough. Users are only being taught about searching, but will spontaneously evolve browsing strategies to deal with problems of recall and precision. If this interpretation of the evidence is correct, then the tutor should expand their information retrieval syllabus to give browsing strategies equal status with searching strategies, and should give advice on solving retrieval problems with both techniques. If the findings of the GAELS Project are representative, user education practice does not accept such strategies as integral parts of the user education syllabus, but should do so.

If our investigation of user education practice is correct, then British user education practice in fact differs from important statements of library user education theory, such as the Model Statement of Objectives in library user education [17] by the Association of College and Research Libraries of the United States. The ACRL outlines, as a terminal objective of bibliographic instruction, that the 'user understands how...to identify useful information' by knowing 'when it is appropriate to search for information', but also 'understands the importance of browsing'. Our argument that browsing deserves its rightful place in the user education syllabus is thus only a statement of accepted library educational theory.

This suggestion that librarian teachers accept users' task behaviour and consider redesigning their systems (in this case, their user education syllabus) to accommodate such behaviour is an implementation of task analysis ideas for interface design. As Lewis and Rieman say, 'Once you have some people to talk with, develop CONCRETE, DETAILED EXAMPLES of tasks they perform or want to perform that your system should support.' When you find that people perform different tasks from what you originally thought, do not be 'shocked' (Dyckman's phrase) at your discovery. Rather, congratulate yourself and re-design your product to accommodate your improved knowledge of the user.

Ironically, our suggestion for syllabus redesign could also be construed as adding a third argument to the anti-user education school's repertoire of objections. Not only should librarians stop teaching and improve the usability of their libraries, leaving lecturers to teach information skills in the context of coursework. They should also stop teaching inappropriately complex search techniques to users, and should leave them to augment basic search skills with other browsing skills, which can be evolved more effectively in the context of departmental teaching and research. For example, the highly information literate technique of reference chaining from colleagues' bibliographies can only take place within the context of departmental group research. However, this would be a misinterpretation of our argument. We are merely offering a third way of improving the quality of free-standing information skills teaching courses, which we believe remain necessary (perhaps even a necessary evil) in spite of all counter arguments for the foreseeable future.

Interim conclusions:

Promote electronic browsing, not searching, for current awareness

The GAELS Project was set up to encourage electronic information retrieval in place of hard copy journal browsing. Our initial findings indicate that one way to achieve this culture change may be to stop trying to impose electronic searching techniques on to an established pattern of hard copy browsing. Rather, the virtual information retrieval syllabus should look to methods of electronic browsing, and use these as the desired model of new information retrieval activity among researchers.

In practical terms of syllabus content, this has some simple consequences. Many database services offer a 'save search strategy' option, which allows a one-off literature search to become a stored search to be regularly repeated as a current awareness tool. Rather than recommend this as the first option for a current awareness strategy, a teacher librarian may more successfully promote electronic alternatives to hard copy browsing by recommending electronic table of contents browsing through publishers' own websites. Ultimately, a researcher may feel happier title browsing through such tables of contents of their favourite journals This is especially true if the sites offer electronic full-text retrieval.

Even without electronic full-text retrieval, there are various other reasons for such a user preference A publisher's WWW site often has graphic design features reproducing the look of the hard copy equivalent. These features make it possible to keep some of the feel and familiarity of the library browsing experience in an electronic environment. Users can bring such tables of contents together in a cross-publisher set of WWW pages, creating their own researcher-specific WWW environment which is fully under their own control. In the absence of immediately readable electronic full-text, the option of using hard copy commercial document delivery from scholarly publisher services, in preference to local library document delivery services, may also prove attractive Table of contents browsing can be more thorough than saving and re-running a bibliographic database search, given that much database journal article indexing is selective. Part of the user preference for browsing to reduce recall is item by item control of the article de-selection process, a control that users are reluctant to cede to a machine process which hides the imposition of unknown factors, such as selective indexing. If they browse, they will not miss anything, if they search, they might.

This model of electronic browsing does not use librarians' own preferred services, such as bibliographic databases services purchased with library funds which demand good search skills, and which are normally promoted in tandem with a local library inter-library loan or document delivery service. However, this model may better resemble real hard copy browsing activity and thus may be a model that users actually prefer. Feedback from researchers to the GAELS Project does indeed indicate a valuable role for publisher's WWW sites, rather than library bibliographic databases, in current awareness. We have also found an interest in commercial document delivery where there is a perception that such services offer advantages over the local library document delivery service, above all in terms of delivery speed. These are further examples of evaluative user feedback informing the task analysis process, in order to create a better user education syllabus and thus better forms of user service.

Build other methods into literature searching syllabus

One way forward for the GAELS Project courseware is to build this model into the CAL package and test the acceptability of the approach with users. If an increased emphasis on electronic browsing should be introduced into the current awareness sections of the package, then similarly the sections on general search skills may well need re-balancing in favour of other search strategies. The traditional approach is to teach basic search skills and then introduce advanced search skills to raise or lower recall in an initial set of hits. Users might respond better if asked to adjust recall with a variety of strategies. The users themselves could be invited to try a variety of techniques, but left to judge for themselves which strategy they find more successful.

Emphasise objects retrieved more than techniques of retrieval

Beyond this, further developments to the syllabus also remain possible. Although user feedback has suggested extending and varying the task content of the first GAELS Module, users have expressed satisfaction with the content of the Modules on Information Tools and Information Types, implying that they unambiguously need help with knowing where to look for information and advice on what they find when they look there. The sequence in which the GAELS Modules are presented (which reflects existing local teacher librarian preferences) may thus be inappropriate. A re-working of this presentation sequence, in favour of the types and tools of information retrieval, might further help in countering an over-emphasis on how to do information retrieval in certain prescribed ways. Again, more evaluative user feedback would show us whether this alteration was an enhancement or not.

Use Library WWW pages as task-based interface to courseware

Perhaps most importantly, a significant development of the task-centred focus of the courseware would be to provide a task-oriented interface to the package by plugging appropriate collections of units into local library WWW pages. These collections would each be equivalent to 'real, complete, representative tasks'. This would necessitate a prior re-working of the local library interface itself to display a significant degree of task-orientation. But in this way the linear narrative structure of the original package interface could co-exist with a more usable interface, so satisfying teacher librarian need for a classroom, or workshop teaching tool, at the same time as meeting user need for training in a particular skill at the point of execution of a task. Thus, a researcher needing to complete a literature search would go to their local library WWW pages and on a top-level menu find an option flagged 'Do a literature search'. If they lacked the skills to do this, they would have the option of completing the relevant element of the GAELS courseware at this point.

Use departmental courseware as task-based interface to courseware

Taking this a step further, a more complete integration of learning materials into task execution would be possible by using WWW-based departmental teaching and research training materials as, in a sense, the interface to our library courseware. Thus, at particularly appropriate points in a piece of departmental courseware, where an information retrieval activity was required, the local library's information resource links and the courseware supporting the use of these links could be invoked. If such a scenario was carried to a conclusion, user education could theoretically cease to exist on a separate free-standing basis, being completely integrated into the teaching of a department.

If it were possible to use departmental courseware as the preferred task-centred context for library courseware, we would have moved full circle, back to the original aims of Herrington and similar writers: library interface transparency and the reintegration of information skills acquisition into the context of users' mainstream coursework or research. If the final GAELS model emerges as one in which library courseware is taken into the texture of academic teaching and research, via a process of library interface evaluation which first embeds that courseware in the texture of the electronic library service front end, then the primacy of free-standing library information skills user education will have been challenged.

However, this goal will not be achieved without starting from existing, successful user education materials, which can then be continuously developed to their full potential. Even then, the absence of significant provision of WWW-based departmental research and learning environments means that the ideal, contextual framework within which to propagate networked information skills education has not yet been created. Library user education providers will have to wait on such environments developing, and work proactively with academic departments to promote them.

In the meantime, the GAELS Project will explore these hypotheses further and examine whether they can create a more effective tool for the ultimate project end, a joint electronic information environment in engineering that is found acceptable and usable by researchers at both collaborating institutions.

References

1. Barton, J., Ashworth, S. and Joint, N. The GAELS information audit. In. *3rd Northumbria International Conference on Performance Measurement in Libraries and Information Services* held at Longhirst management and Training Centre, Longhirst Hall, Northumberland, England, 27-31 August 1999. Newcastle Upon Tyne: Information North for the School of Information Studies, University of Northumbria, 2000, 139–43

2. Pacey, P. Teaching user education, learning information skills; or, towards the self-explanatory library. *The New Review of Hypermedia and Multimedia*, 1, 1995, 95–103.

3. Savenije, B. Integrating library instruction into student education. *Education Libraries Journal* 42(1), 1999, 5–10.

4. Herrington, V.J. Way beyond BI: a look to the future (the effect of technology on bibliographic instruction). *Journal of Academic Librarianship*, 24(5), 1998, 381-386.

5. Rutkowski, C An introduction to the human applications standard computer interface. Part. 1. *Byte*, 7(10), 1982, 291-310.

6. Shneiderman, B. *Designing the user interface: strategies for effective human-computer interaction.* 3rd edition. USA: Addison Wesley Longman Inc., 1998.

7. Snavely, J. and Cooper, N. The information literacy debate. *Journal of Academic Librarianship*, 23(1), 1997, 9–14.

8. Bober, C., Poulin, S. and Vileno, L. Evaluating library instruction in academic libraries. a critical review of the literature 1980–1993. *The Reference Librarian*, 51/51, 1995, 53–71.

9. Lewis, C. and Rieman, J. Getting to know users and their tasks. In: *Task-centred user interface design: a practical introduction.* Chapter 2. (self-published over the Internet· 1993) In: Baecker, R.M , Grudin, J., Buxton W.A.S. and Greenberg, S. *Readings in human-computer interaction: toward the year 2000.* 2nd edition. San Fransisco: Morgan Kaufmann, 1995.

10. Strathclyde University Library. About the Library WWW pages. http://www.lib.strath.ac.uk/Strath_Lib.

11. EDUCATE Project's *Into Info* courseware. http.//educate.lib.chalmers.se/demopath.html

12. GAELS courseware. http://gaels.lib gla.ac uk/info_skills/.

13 Kemp, R Training materials for advanced information skills in engineering and architecture: a web-based approach to advanced information skills. In: *Building on the past: Proceedings of the 12th ARCLIB International Conference, University of Strathclyde and Glasgow School of Art, 1999.* 10–16.

14 Bates, M.J. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 1989, 407–424.

15. Marchionini, G. *Information seeking in electronic environments.* Cambridge Cambridge University Press, 1995.

16. Dyckman, L M Beyond 'First you push this button, then...': a process-oriented approach to teaching searching skills. *The Reference Librarian*, 51/52, 1995, 249–265.

17. Association of College & Research Libraries. Model statement of objectives for academic bibliographic instruction. draft revision. *College & Research Libraries News*, 48(5), 1987, 256–261.

Subscriptions
Aslib Proceedings is published ten times per annum and supplied to corporate members of Aslib. For 2000 the annual subscription rates are £170/$265 (£136/$212 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 20 7903 0000. Fax: +44 (0) 20 7903 0011. Email: aslib@aslib.com *http://www.aslib.com/*

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
POSTMASTER· Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

© 2000 Aslib and contributors ISSN 0001 253X

0001-253X (2000) 52:8 ;271-312

ISSN 0001 253X
Volume 52
Number 9
October 2000

Aslib Proceedings

New information perspectives

Digital health information: case study the information kiosk
David Nicholas, Peter Williams and Paul Huntington

Screen-based information design: Eforms
Alistair Ross

A definition of information
A.D. Madden

What should we understand by information technology (and some hints at other issues)?
Julian Warner

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

Aslib Proceedings is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In some cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000–8,000 words in length. Articles should be accompanied by a 100–150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The editor is happy to provide advice on an informal basis to potential authors.

Editorial Board

Dr David Nicholas, *Head of Department, Department of Information Science, City University*; John Akeroyd, Director, *Learning Resources, South Bank University*; Sarah Blair, *Head of Publications, Aslib*; Peter Chapman, *Information Manager, NewQuest North East*; Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire*; Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw*; Harry East, *Senior Research Fellow, Department of Information Science, City University*; Mike Hyman, *Managing Director, Head Software International*; Robin Hunt, *Creative Director, AreHause*; Nat Lievesley, *Network Manager, Centre for Policy on Ageing*; Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Senior Lecturer, Department of Information Science, City University*; Richard Withey, *Director, Independent Digital*; Ralph Adam, *Freelance Journalist;* Professor David Ellis, *Department of Information and Library Studies, University of Wales;* Professor Ben Fouche, *Stellenbosch University*.

Editorial Team

Editor: David Nicholas
Deputy Editor: Ian Rowlands
Editorial Manager: Kay Nicholas
Editorial Assistants: Peter Williams (City University), Colette Coles (London Research Centre), Paul Huntington (City University)

CONTENTS

315–330 Digital health information: case study the information kiosk
David Nicholas, Peter Williams and Paul Huntington

331–342 Screen-based information design: Eforms
Alistair Ross

343–349 A definition of information
A.D. Madden

350–370 What should we understand by information technology (and some hints at other issues)?
Julian Warner

Digital health information: case study the information kiosk

David Nicholas, Peter Williams and Paul Huntington
Internet Studies Research Group, Department of Information Science, City University, London EC1V 0HB
nicky@soi.city.ac.uk
http://www-digitalhealth.soi.city.ac.uk/isrg/doh.htm

The published literature on consumer health kiosks is assessed to determine the impact of this relatively 'low-tech' technology and to provide a context and stepping-stone for a Department of Health (DoH) funded study, which is evaluating the impact of more than 70 health kiosks located in all kinds of locations throughout the United kingdom[1]. The existing literature indicates that kiosks are benefiting health information provision in some important ways. Data from some early pilot work on the DoH project illustrates the kinds of analyses that can be conducted and appears to confirm the popularity of the kiosks with the general public.

Introduction

While everybody has been watching for the latest Internet development – and the WAP mobile phone connected to the Internet has particularly caught the eye – touch-screen digital health information kiosks (and their hybrid forms) have quietly spread themselves around the country. There are probably around 200 of them altogether located in surgeries, hospitals, health centres, shopping centres and even airports and railway stations. This number is likely to double over the next couple of years. They are not high-tech or particularly high profile, but maybe they are just the technology needed – relatively cheap, robust, (supposedly) accessible to all and easy to use. The Internet Studies Research Group at City University are currently undertaking a Department of Health funded study 'Digital Health', which is examining the impact of these kiosks on the health consumer and comparing them with other health information platforms, such as the Internet and digital, interactive television.

As part of this research the use and influence of nearly 80 touch-screen kiosks provided by In Touch With Health will be examined, largely through an evaluation of their use logs. The daily logs from these kiosks enable us to determine:

- The volume and pattern of use – hourly, daily, weekly and monthly data can be obtained. Use can be measured in terms of both pages viewed and amount of time spent viewing the pages. Some of the data goes back more than two years, so the identification of trends should be possible.
- The characteristics of use – particular health pages viewed, and whether that information is printed or not.
- The characteristics of the users – number of users, their age and gender.
- The differences in use patterns between kiosks in various organisations (surgeries, hospitals etc) and those located in different parts of the country – the kiosks are spread through inner city, suburban and rural areas in the UK.

The postcodes of the kiosks will also enable us to relate the data to census and medical information and help answer such questions as whether people in an area subject to certain types of illnesses reflect this in their information seeking. These quantitative methods will be supplemented by a comprehensive qualitative analysis from in-depth interviews with information consumers such as patients, doctors, carers etc.

---

[1] The Web, the kiosk, digital TV and the changing face of consumer health information provision: a national impact study. April 2000 – September 2001.

Existing research

Early work

Surprisingly for a technology that has been with us more than a decade now kiosks have not really been the subject of substantial academic research. Very few studies appear to have examined health information mediated in this way, despite government initiatives to install them in a number of public places. Of those that have been undertaken most have been small-scale studies. The first touch-screen health information system was probably Healthpoint, developed by Ray Jones and colleagues at Glasgow University [1, 2]. Indeed, Jones and his team at the University's Centre for Consumer Health Informatics are responsible for the lion's share of research on the topic. Their first prototype system, developed in 1988–89, used a keyboard where nine coloured keys represented menu items. It was placed in a 'very public place' at the front of the waiting area of a health centre [1]. Information was included on drugs, contraception, AIDS and other health topics, and there were also lists of local facilities and groups. Thirty-five people were interviewed about the use of the kiosk, almost all of whom were positive about the system. Only 2 (6%) thought Healthpoint was not easy to use, whilst 32 (91%) were in favour of retaining the facility on a more permanent basis.

A questionnaire survey of the users of the health centre was also undertaken – 151 responded. Thirty-four per cent said they used Healthpoint. Use was not related to sex, age, neighbourhood type, or length of time spent in the waiting room. However, it appeared that 63% of respondents used the kiosk simply out of curiosity, with only 37% seeking specific information. Of the latter, 95% found the information they were looking for. Plainly, mixed signals here.

The first touch-screen version of Healthpoint appeared in 1991. Twenty-five terminals were sited in a variety of places in Clydebank, including a chemist, health centre, library and a public house. Ten were formally evaluated. A random telephone survey undertaken five months after installation showed that 17% of the sample had used at least one of the terminals. The machines themselves were monitored, and the most accessed pages were *Smoking, AIDS, Alcohol, Women's Health* and *Sex* [3]. A later study of Healthpoint [4] showed, a little confusingly, that although only 65 search 'episodes' were logged, the system was used by a total of 116 users. This anomaly is explained by the fact that if a second user takes over from someone before the system times out, only one search session is registered. 'Real' user numbers, and not those derived from logs, were only established by the use of a video camera.

The most popular sites continued to be *Smoking, AIDS, Alcohol, etc.* as in previous studies, but *Stress* moved up in the rankings, and *AIDS* dropped a little – perhaps reflecting the differing anxieties of society as the 1990s unfolded. Evidence was obtained which showed that people had a high recall of what they had read on-screen, but why this should be is not clear. A Spanish version of the system has been developed; the evaluation of which is in press [5].

Influential factors

Pearson *et al* [6], also working with Jones, outlines a number of factors that may influence the accessibility of information systems for the general public. These include:

- Faith in the content (i.e. the question of authenticity) – this might have an impact on kiosks located in shops as opposed to those in surgeries or health centres.
- Ease of use – the authors point out that both hardware, navigation and other aspects of the interface are important and that what constitutes ease of use may be different for naive rather than regular computer users. They cite Al-Barwani [7] who found that only 20% of 200 gastroenterology patients had used a computer, and that 75% of the frequent kiosk users still chose the touch-screen option.
- Convenient location at a convenient time –

interestingly, the authors point out from their own previous studies [8] that this might not necessarily be at the doctor's surgery or in a hospital department, despite the fact that these are the places where 'patients' will always need to go. This is because 'many [visit the doctor or hospital] in the car of relatives or friends, or possibly hospital transport, and feel the pressure of time as well as the general anxiety of their visit'.

- Cost – plainly an attraction of kiosks is that they are free to use.

The authors report on the accessibility of two similar touch-screen computer systems, used in a randomised trial in Scotland, by patients with cancer. Use, repeat use and ease of use of the system are reported and compared to demographic characteristics, level of anxiety, prior computer experience, attitudes and information need (as expressed by the patients themselves). The study was undertaken at The Beatson Oncology Centre (BOC) in Glasgow, which provides cancer treatment serving the West of Scotland. Two computers were available for use by patients; one for use in the researchers' office, and a purpose built booth, to facilitate an element of privacy, in a nearby waiting area at BOC. Both computers had 17" touch-screens, with the interface designed to be simpler to use than a web browser. Patients were those who knew they had cancer, and were to receive radiotherapy treatment. Of the 345 subjects chosen, 167 used a 'general' system that did not include any customised or personal information about themselves, 85 a system that did so and started from these, and 93 a system in which patients were able to choose whether to start with personal or general details.

Data were collected on anxiety and depression (HADS), expressed information needs, previous use of technology and newspapers read. Post-test opinions were sought on ease of use and relevance and utility of the information. A questionnaire was also completed at home three months later about the use of the printed information and patient preferences regarding time with a professional as compared to time with a computer.

Nineteen per cent agreed that using the computer made them feel a little anxious, but 50% had said they were looking forward to it, and 32% were neither particularly interested nor anxious. Fifty-five per cent had never used a computer before and of those who had 39% had only used one a few times. Eighty-four per cent wanted as much information as possible, with the rest either wanting good or no news.

The patients' attitude towards obtaining information was the best predictor of the length of time spent on the computer. Although not highlighted in the article, this presumably means that information needs outweighed computer competence or familiarity. Seventy-eight per cent found using the computer easy or very easy, so good IT skills may not have been required anyway. However, the numbers reusing the computer 'were not great'. One explanation for this might have been the poor location within the hospital. Perhaps also their needs had been met.

Research has also been conducted into the use of a touch-sensitive video monitor for entering health-related quality of life. In Buxton, White and Osoba's [9] research, 178 patients answered a 'Quality of Life'[2] questionnaire using such a system. They were then interviewed on their experience of using it. They were asked, in particular, questions about their views on its ease of use, in terms of entering an ID number, reading and using the screen and following instructions. Results showed that only '1–2%' of subjects found any aspect of the system 'somewhat difficult', and none at all rated any aspect of the system 'very difficult'. 'The majority' found all aspects 'very easy' to use, and the time taken to complete the questionnaire was acceptable to all (mean 6.8 minutes) and 98% indicated that they would be willing to complete a similar questionnaire in the future. The authors conclude that the touch-screen system appears to be 'a highly acceptable approach' for the collection of this kind of data in clinical practice.

---

[2] The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire, the QLC-30.

Children's use of kiosks

Douglas, Jones and Naven [10] extended their research into the impact of touch-screens to look at children's use of such a system. Two studies were undertaken. In the first, pupils at a Catholic school in Glasgow were invited to look at a Healthpoint touch-screen kiosk and produce 'stories' for new topics, which they later developed using a computer based presentation package. Topics the children suggested tended to be oriented around the general themes of sex and eating, and included 'stories' on teenage pregnancy, rape, abortion, anorexia and bulimia. Dyslexia was also a chosen subject. Some of these issues were not originally included on Healthpoint, and were later incorporated as a direct result of this exercise. Thus, pupils saw their ideas rapidly being disseminated in the wider community. The authors also point out that the IT aspect of the exercise may have given health education a certain 'street cred', which it may have lacked with a more traditional (poster and leaflet) approach.

In the second study, three schools were given terminals and pupils and staff were interviewed about their experiences in using them. The two Catholic schools participating requested censoring of the information. The whole section on sex was removed (the authors point out that this included contraception and sexually transmitted diseases), a reference to sex in the AIDS section, all references to drugs as being 'pleasurable' and other material. Given these omissions, the most popular topics were 'women's health, AIDS and smoking. At the other school sex was the most popular topic – possibly as feared by the censoring institutions.

In feedback questionnaires and group discussions, girls tended to complain that the system should have been more private. The closed nature of the community, where anonymity is impossible, appeared to be a major factor in this recommendation. In general, pupils wanted more information than that offered, possibly an oblique reference to sex, although some gave specified drug related information as being missing. This included details of 'what actually happens to people who've had bad experiences with drugs'. High achievers said the system did not tell them anything new at all. The conflict between the interests and aspirations of the pupils and school management (as represented both by the censorial teachers and parents) was also recognised as a problem by the researchers. However, the hope is that by the long term installation of a health information kiosk and its constant updating and evolution, the basis could be established for permanent and dynamic health promotion 'rather than the usual *ad hoc* campaigns'.

Comparison between the use of digital and paper sources

In their study, comparing media for patient use, Graham *et al* [11] looked at the effectiveness of a touch-screen system with an information leaflet for providing women with information on prenatal tests. Two factors led to the choice of this subject: it is an area in which informed choice for patients has 'long been applied', and technical and medical advances have meant that the number of conditions for which screening is available 'continues to grow rapidly'. Perhaps for these reasons, antenatal screening 'is one of the most intensively researched topics with regard to information for women'.

The principal measure the researchers were interested in was the effect on the patients' decision making, as measured by their understanding and take up of the tests. Other outcomes included their satisfaction with the information provision and their anxiety levels. Their understanding was measured by pre-, during and post test questionnaires, and anxiety by the Spielberger state-trait anxiety inventory, which was included in the questionnaires. Data on their use of each information source and satisfaction levels appears to have been elicited by interview.

The sample was obtained from women attending antenatal clinics at Aberdeen Maternity Hospital. One thousand and fifty consenting women began the experiment, with 875 continuing until the end. Subjects were randomised on a 1:1 ratio into an intervention and a control group. Both groups were given an information leaflet developed for the research, and the intervention group also had access to the touch-screen system. The information on the leaflet was included in the electronic system, but the latter also went into greater detail and had a wider scope. Plainly not a level information playing field, but one would naturally expect more information to be provided electronically.

Results showed that similar numbers of women from both groups read the leaflet 'fully' (57% intervention group, 61% control group) and 12% of both groups 'glanced at it'. Ninety-one per cent of the intervention group indicated that they had used the touch-screen 'at least once'. Women's understanding of tests 'improved significantly' for both groups during the course of the project, but there was no difference in performance of the touch-screen over the leaflets in informing the subjects. However [12], medical students at the University of Newcastle point out that the questionnaire only assessed knowledge of the purpose of the tests and not the risks/benefits, so the level of understanding could only be measured in these terms.

With regard to take up of prenatal tests resulting from the information provision, only 1% of women did not undertake any of these. However, of five offered, only one – an anomaly scan – was undertaken by a significantly different number of women from the intervention group: 97% versus 87%. The authors suggest that the use of video clips to show what can be gained from such a scan 'might have helped reassure women and increased their desire for this investigation'. However, regarding take up of tests as an indicator of 'informed decision making' the research has been criticised, again by Williams *et al* [12]. These commentators consider that 'informed women may decide not to undergo testing for reasons other than a lack of understanding of the procedures'.

Satisfaction with both media forms was high, with 95% of both groups stating they would recommend the leaflet, and a similar percentage of the intervention group said they would recommend the touch-screen. 'Over one third' indicated a preference for the screen over the leaflet, with 'a fifth' opting for the leaflet and a quarter indicating no preference.

Interestingly, the first comment of the researchers regarding anxiety levels was that these were not *raised* by the information provision (!) In fact, the level for those using the touch-screen was actually reduced, whilst there was no significant difference in levels before and after receiving the information for the leaflet only (control) group. Disappointingly, the possible reasons for this are not explored. Jones and Craig [13] suggest that had the sample been drawn from a less affluent catchment area, 'where non-print media may have benefits over the written word among a population with lower literacy and knowledge levels', the benefits of the touch-screen delivery may have been greater, resulting in even better results in terms of anxiety level reduction and knowledge acquisition.

The study concluded that 'the touch-screen evaluated in this trial conferred no additional benefit to that provided by ... the information leaflet'. Nevertheless, some advantages were noted. The increase in anomaly scan take up in the intervention group, the level of preference for the screen, and the reduction in anxiety all indicate the value of the touch-screen as an effective medium. The authors do, however, imply that the touch-screen is expensive – the system used cost £25,000 and 'additional costs can be envisaged in terms of the maintenance of hardware and updating the system'. Jones and Craig [14], however, argue that the researchers did not consider the transferability of the system to other locations and the consequent potential high take up. They calculate that the cost per user could be as low as 26 pence, cheaper than buying and giving access to 'expensive (typically £3) cancer booklets'.

Jones [15] and colleagues undertook their own comparative study of information delivery systems. They compared the use and effect of three information sources for cancer patients: a personalised computer based information system, a general computer system and booklets. They were particularly interested in patients' views and preferences, use of computer information and psychological status, and relative cost. They argued that 'most cancer patients want as much information as possible and wish to be involved in treatment decisions'.

The sample was of patients suffering from breast, cervical, prostrate laryngeal cancer. Those suffering severe pain, or who had a visual or mental handicap, were excluded. The resulting 525 subjects were divided into three intervention groups, according to the information system they would be offered (personalised computer, general computer or booklet). Half of the group receiving personalised information also had access to the general computer information archive.

Unlike in the study by Graham *et al* [11] the computerised information systems automatically recorded time online and pages consulted. For the booklet information group, the number and type of booklets chosen was manually recorded. Patients' views on the information were gathered from questionnaires. Anxiety and depression were measured before, during and after the intervention by standard clinical methods. Costs were calculated to take into account service requirements (research assistant salary) and maintenance charges, but not capital costs. Costing the booklets was somewhat easier, being simply the costs of the booklets taken by each patient (as paid by the researchers).

The results of the study showed that the average time per computer use session was 12 minutes (range: 1–44 minutes). Twenty-nine per cent of patients used the computer a second time during a three-month follow-up period. Of those who had access to both personal and general information, 'two-thirds' went to the personal information section first. Booklet information use was described purely in terms of home use. More patients (83%) used the material at home compared with those offered personal (70%) or general information from a computer (57%), printouts of which were made available for home use. This was partly attributable to the booklets being described as 'more attractive than computer printouts'. Interestingly, more of the patients offered booklets felt overloaded with the information. There were a large number from which to choose, and the computer groups did not appear to spend long enough on the systems to appreciate the amount of material housed on them – they only printed out relevant information and were consequently in control of the amount of information they received.

As regards cost, the researchers calculate that it would cost 900% that of booklets to produce a personalised computensed information service for patients, a figure exacerbated by the absence of electronic patient records. However, the authors contend that this would fall to the same price as a general system once patient records were online 'as fewer additions would be required'. This worked out at 40% of paper based information provision cost (assuming full access to all the books). For comparisons, the cost of the booklets averaged at £7 per patient.

There were no significant changes in the patients' depression scores or 'mental adjustment' scores during the length of the experiment. However, there was some evidence that the generalised computer information group 'were more anxious than other patients' with their scores, although declining during the course of the experiment, rising to pre-test levels by the end on a tabulated 'hospital anxiety and depression scale'. This also showed the personal computer and booklet group scores lowering as time went on, indicating that, for them, the information had some element of reassurance.

Velikova *et al* [16] compared a touch-screen system used to complete a 'Quality of Life' questionnaire, and also a Hospital Anxiety and Depression Scale, with paper based equivalents. One hundred and forty-nine cancer patients were selected, a similar sample size to that of Buxton [9], all of whom completed the questionnaire in both formats. Fifty-two per cent of patients preferred the touch-screen, 24% the paper version, and the remainder (24%) had no preference. Patients appeared to find the touch-screen simple to use, and there were no missed responses. There was also near equivalence in scores for the paper and the touch-screen versions, although patients 'tended to give more positive responses' to emotional, fatigue, appetite and nausea questions. The researchers also tested the reliability of the touch-screen system by asking 81 other patients to complete the questionnaires electronically twice, with an interval of three hours between tests. Correlation co-efficients and other measures indicated that the questionnaire 'had a good test-retest reliability'. The authors conclude that touch-screen questionnaires were well-received and accepted by cancer patients, and had good data quality and reliability.

Comparison between use by health professionals and lay people

Patyk *et al* [17] looked at how touch-screen systems could be used in patient education. They measured the attitudes and perceptions of medical staff and lay people on the effectiveness of a multimedia computerised package operated by a touch-screen system – the Brain Injury Resource Center (BIRC). This is a multimedia CD-ROM learning centre offering 23 hours of educational information on traumatic brain injury. The system supplies basic information, defines medical terms, and allows users to see and hear real patients and their families tell about their brain injury experiences. The BIRC rests on its own base, can be viewed easily from either a standing or sitting position, and is wheelchair accessible.

Data was gathered from a written survey of 31 healthcare professionals and a thirty-minute interview of patients with brain injury and their families. Both questionnaires and interviews covered the same ground: the usefulness of the information, the effectiveness of the teaching methodology, and the most appropriate time frame within the care process for implementation. Both the medical staff and patient participants viewed the BIRC for at least 15 minutes, although most spent more than the required viewing time. Results from the survey suggest that participants most chose to view:

- 'Commonly asked questions' (52%)
- 'The Injured Brain' (45%)
- 'Treatment and Rehabilitation' (45%)

Nurses (half the sample) were interested in 'Treatment and Rehabilitation', 'Preventing Brain Injury' and 'Resources'. This group also felt strongly about accompanying their patients during the introduction to the BIRC. The features that respondents liked best included viewer control (71%) and graphics/illustrations (58%). More than 75% of the respondents felt the BIRC was effective and non-threatening and that it allowed for control of the pace and content of the material. More than 50% of those surveyed rated the information presented as being highly understandable (65%), useful (58%) and important (55%).

In the patient interviews, the wealth of information, readability, glossary and the presentation of real-life cases were all praised. Interviewees said that the touch-screen methodology, format and instructions made it very user-friendly, and that it was 'easy to understand' even though they were not computer literate. Patients were also asked to identify an appropriate time for introducing family members and patients to the BIRC, but there was no consensus on this topic. It was felt that the timing would vary with each person and situation, and that the resource should be made available during all phases of care.

The benefits of the BIRC were perceived differently by the healthcare professionals and patients/families. All the latter selected 'to provide background information' as a benefit of the BIRC, whereas only 32% of healthcare professionals did so. Both groups were similar in their assessments of the system's ability to stimulate questions and reinforce 'verbal information'. However, patients' and family members also selected 'calm fears' and 'primary means for attaining information' as benefits, whereas healthcare professionals rated these benefits significantly lower.

Another discrepancy noted was that healthcare professionals tended to view computer-based education as supplementary or peripheral to their direct hands-on instruction, whereas patients and their families viewed it as more central to their quest for information.

The authors conclude that 'computer methodology', such as that used in the BIRC, can play an important role in meeting the educational needs of patients with brain injury and their families. It was felt that such systems could:

- provide the necessary background information;
- define the needed medical terminology;

- prepare users (via video vignettes) for some of the expected behaviours associated with brain injury, including those that involve speech, gait and emotional responses;
- enhance educational sessions by incorporating a multisensory approach.

The authors maintain that it is most valuable, however, when it supplements human interaction with healthcare professionals.

Related work

Doctor interaction with information is beyond the scope of this paper. However, one study [18] does appear to be particularly relevant, because it examines how the format of information presented can influence clinical decision-making. Although the study was exclusively concerned with clinical decisions made by doctors, there are clear implications for the design and presentation of health information for consumers too. Study participants looked at four displays of hypothetical trial data from two clinical treatments, one 'conventional' and the other 'investigational', and were required to say whether they would continue the trial or stop it for statistical analysis. One of the treatments was 'clearly superior to the other' and 'because of the large difference in efficacy, it would be clinically appropriate ... to stop the trial'.

The displays shown were tables (text based), pie charts, bar graphs (stacked) and icons (similar to pictogram format) In each case data was presented from the same hypothetical trials, although none of the participants commented on the similarities, and one went so far as to complain that the data were so different that display comparisons were 'meaningless'. For each display, data was recorded on the decision taken and time required on deciding. Display preferences and other comments were also noted.

Results showed that mean times for decision making were very similar for each display, and was not a function of the success rate of responses. There were, however, variations in decisions made according to the type of display considered. The icon representation produced the most correct decisions (82%), with the pie and bar charts having a success rate of only 56%. The tables were framed in two ways, negatively and positively. The former displayed failure rates for the treatment, and the latter success rates. This framing resulted in a big difference in decision making. Only 47% of decisions were correct based on the negative representation, compared to 93% for the positive one (overall correct decisions for the table format were 68%).

Interestingly, preferences as regards data display were not a factor in the decisions made. The most popular display, the table (62% of subjects), yielded only 68% of correct decisions (the second best out of four), whilst the least popular, the icon, which none of the subjects preferred, actually scored the best in terms of correct decision making. This was put down to its attribute of showing every treatment response individually. There was, however, no difference in success rate and order of data representations.

Reasons why medical professionals made incorrect decisions, discussed by the authors, is beyond the scope of the present study[3], but their general conclusions are. These were that the method of displaying data and the framing of the information 'significantly influenced' the accuracy of decisions. Of significance in terms of touch-screen displays for health consumers – the lay public – is that the research implied that the method by which health information is displayed generally could lead to incorrect decisions. This may be particularly true with regard to health risk data. Government or official figures may be presented in such a way on screen (or, indeed, on paper) that readers may feel unduly alarmed or base decisions on misinterpretation of data. On the other hand, it may be convenient for the government to present information in a deliberately negative or positive light, to 'engineer' a desired outcome. Information on smoking related disease may be negatively framed, whilst that on CJD presented in a way that minimises possible dangers. These are real issues in the field of consumer health information that need to be addressed by both providers and mediators.

---

[3] Reasons include the hypothetical nature of the trial, reducing subjects' care and attention and a disinclination to stop a trial after a single observation. The authors note that, using this reasoning 'erroneous decisions in (the) experiment might not be considered errors by some investigators'.

The City University health kiosk logs pilot study

A pilot study was undertaken of the logs of four recently installed kiosks supplied by InTouch With Health; they were located in hospitals in Truro, Penzance, Oxford and Stockport. The period covered by the pilot survey was September 1999 to April 2000, although individual kiosks did not contribute data for the whole period (Table 1). During this period 8,685 people used the kiosks and they viewed 82,442 pages (Table 2). By any standards these are impressive figures. Looking at the monthly figures (Table 3) Truro proved to be the busiest site recording nearly four times the use of Penzance in February. An average of 9.5 pages were viewed per person. This varied a little, from 8.8 in Stockport to 10.9 in Oxford.

Table 1. *Time span of data*

Penzance	January to April 2000
Truro	January to April 2000
Oxford	February and March 2000
Stockport	August 1999 to March 2000

Table 2. *Numbers of users and pages viewed*

	Number of Users	**Amount of Use**	**Use per user**
Truro	4269	40523	9.50
Oxford	1413	15358	10.90
Penzance	1792	15867	8.90
Stockport	1211	10694	8.80
Total	8685	82442	9.50

Use proved to be very volatile. Figure 1 shows this very clearly for Truro. In this case use has fallen over the four-month period surveyed – maybe as a result of the moving of the kiosk around the building. Day of the week would seem to be an influential factor. Thus of the four highest peaks in demand, three occurred on a weekend. A Friday (24th March 2000) saw use drop to its lowest figure, and another Friday (25th February) also witnessed a big fall. Peak use in fact occurred on a Sunday (30th January 2000). However, the relationship is not always a strong one. Thus 30th March saw use drop considerably – and that was a Sunday.

Table 3 *Page use by month and place*

	Place			
	Truro	Oxford	Penzance	Stockport
September				2048
October				1846
November				1810
December				907
January	9108		5116	1098
February	16082	10490	4809	1407
March	12440	4868	4129	1578
April	2893		1813	
Total	40523	15358	15867	10694

Characteristics of users

On beginning a search users are asked to click on their gender and age group. Interestingly, the elderly, who one might expect to be a significant group in hospitals, were not big users of kiosks (Table 4). Just 18.0% were over 56 years old. If illness is not the chief determinant of use then perhaps IT literacy is – 22.6% of users were under 15 and 53.3% under 35. Plainly kiosks appeal to the young and the very young, either because they are inquisitive by nature or because they were simply bored and sought to while away the time by 'surfing' for information. There were large differences between locations. Stockport had the youngest users – surprisingly one-third were under 15. The question here is whether they were playing while waiting for the doctor or whether we really do have a cyber generation in the making. Truro and Stockport tend to have the oldest users – 5.2% were over 75, compared to 3.9% in Penzance, although there really was not much in it.

Figure 1. *Pattern of daily page use – Truro*

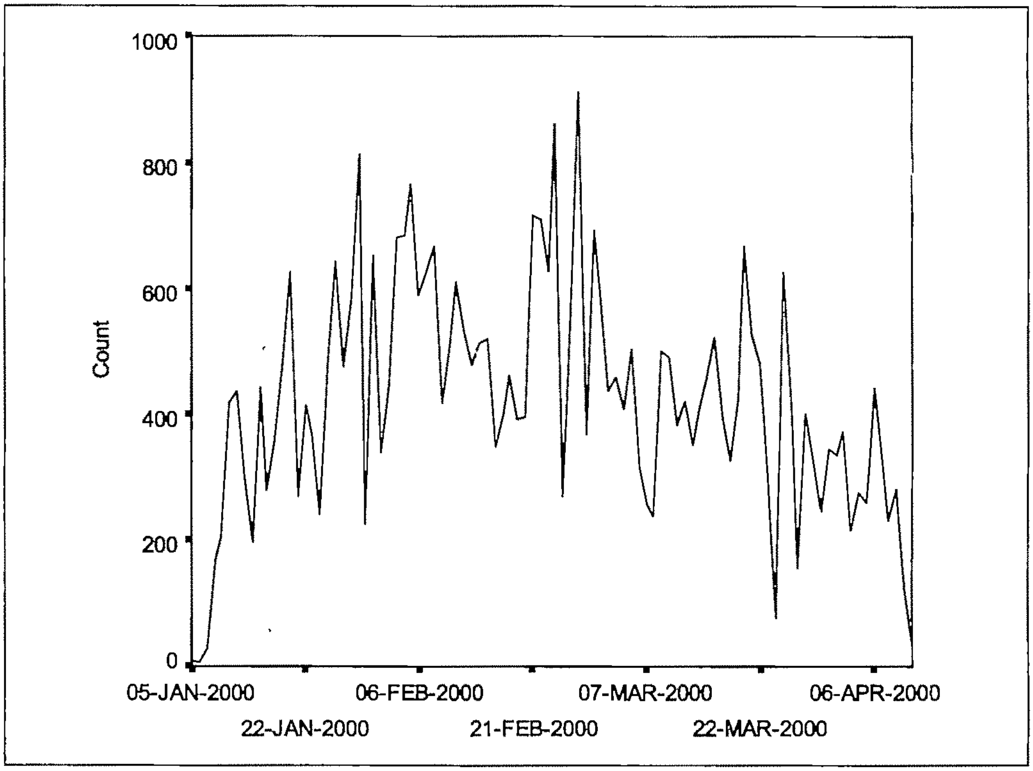


Table 4. *Pages viewed by age of user and place*

	Age					
	Under 15	16 to 35	36 to 55	56 to 75	Over 75	**Total**
Truro	9554 23.6%	11931 29.4%	11286 27.9%	5701 14.1%	2051 5.1%	40523 100%
Oxford	2885 18.8%	6105 39.8%	4406 28.7%	1345 8.8%	617 4.0%	15358 100%
Penzance	2673 16.8%	4404 27.8%	5767 36.3%	2397 15.1%	626 3.9%	15867 100%
Stockport	3507 32.8%	2880 26.9%	2233 20.9%	1521 14.2%	553 5.2%	10694 100%
Total	18619 22.6%	25320 30.7%	23692 28.7%	10964 13.3%	3847 4.7%	82442 100%

Slightly more females (51.6%) than males (48.4%) used the kiosks, but there was really very little difference between the sites with the possible exception of Penzance, where 53.9% of the users were female (Table 5).

Table 5 *Pages viewed by gender and place*

	Gender		Total
	Female	Male	
Truro	20599 50.8%	19924 49.2%	40523 100%
Oxford	7983 52.0%	7375 48.0%	15358 100%
Penzance	8550 53.9%	7317 46.1%	15867 100%
Stockport	5429 50.8%	5265 49.2%	10694 100%
Total	42561 51.6%	39881 48.4%	82442 100%

Characteristics of use

Having entered their age and gender users are lead to a main index page. This has six options:

- Medical conditions;
- Surgical operations;
- Health news;
- Support groups;
- Healthy living;
- Health directory.

A tab indicating 'more options' is situated at the bottom of the screen. This leads to two more entries:

- A-Z of the NHS;
- Travel clinic.

Although the construction and layout of each section is not quite the same, activating any of the main menu items leads to a further list of contents. In some cases, such as 'Medical Conditions', the list – often considerable – can be shortened by various methods. One of these is to touch the relevant part of a body diagram to limit the list to entries related to this. Another facility, offered by the Support Group entry, allows users to key in the first letter of the item they seek. Further content selections are then displayed. Indeed, in some cases five screens are required to arrive at the required information.

The following examples illustrate the set up.

Main menu entry: medical conditions

Second level index: list of conditions (Asthma, for example)
Third level index: questions user might ask (What is it? How does it occur?)
Fourth level: information provision, answering question selected

Main menu entry: surgical operations

Second level index: list of operations and body parts that may require operations (Blood vessel system, for example)
Third level index: list of types of operation (Amputation etc.)
Fourth level: list of details user might require about operation chosen (Body parts involved, the operation etc.)
Fifth level: information provision

Main menu entry: support groups

Second level index: list of topics (Abuse, for example)
Third level index (in this case within same screen, appearing on the left-hand side). Types of information user might require: support groups, books, other material
Fourth level: (also appearing on the same screen, in response to a user option): 'Start Search'.
Fifth level: names of organisations, (i.e. Childline)
Sixth level: information provision i.e. details about organisation selected

Of the first level health categories most interest was shown in medical conditions, which accounted for 41.5% of all enquiries (Table 6). Some way after that came surgical operations (17.2%) and healthy living (16.3%). There were regional variations with more interest in surgical operations shown in Oxford and more interest in travel clinics being shown in Stockport. We can also view this data by gender (Table 7). Concerns were remarkably similar in some instances (e.g. healthy living), but there were some significant differences also (women were more interested in support groups, for instance).

Within these general categories it was possible to identify further individual page use for the healthy living, medical conditions, surgical operations and travel clinic categories. Healthy living was represented by 8 pages on the kiosk, medical conditions by 381 pages and surgical operations by 122 pages.

Table 6 *General health categories viewed by place**

	Place				Total
	Truro	Oxford	Penzance	Stockport	
A to Z of the NHS	1427 5.3%	539 5.2%	659 6.4%	289 4.6%	2914 5.4%
Healthy living	4412 16.5%	1652 15.8%	1709 16.6%	1020 16.1%	8793 16.3%
Medical conditions	11434 42.7%	3993 38.3%	4381 42.5%	2587 40.7%	22395 41.5%
Support groups	2593 9.7%	973 9.3%	1117 10.8%	700 11.0%	5383 10.0%
Surgical operations	4782 17.8%	1919 18.4%	1692 16.4%	901 14.2%	9294 17.2%
Travel clinic	2159 8.1%	1360 13.0%	755 7.3%	852 13.4%	5126 9.5%
Total	26807 100%	10436 100%	10313 100%	6349 100%	53905 100%

*Excluding health news and health directory

Table 7 *General health categories by gender*

	Gender		Total
	Female	Male	
A to Z of the NHS	1818 6.5 %	1096 4.2%	2914 5.4%
Healthy Living	4452 15.9%	4341 16.7%	8793 16.3%
Medical conditions	11887 42.6%	10508 40.4%	22395 41.5%
Support groups	3323 11.9%	2060 7.9%	5383 10.0%
Surgical operations	4308 15.4%	4986 19.2%	9294 17.2%
Travel clinic	2138 7.7%	2988 11.5%	5126 9.5%
Total	27926 100%	25979 100%	53905 100%

Healthy living concerns

Of the healthy living concerns alcohol topped the list. with good eating a close second and exercise third (Table 8). Naven *et al* [4], commenting on a much smaller sample and different geographical region, also found alcohol to be what kiosk users were interested in, but it seems smoking and AIDS has been replaced by healthy eating and exercise. There were large regional differences. Thus weight accounted for 11% of all top ten health pages used in Oxford but only 4% of pages in Stockport; smoking accounted for 13% of all top ten pages used in Truro but about 3.5% in other locations; whilst asthma accounted for almost 9% of all top ten pages used in Truro but just 1.7% in Oxford.

Women plainly have different concerns than men (Table 9). Thus alcohol was not their chief concern, good eating was. Smoking and weight were the two other main areas that women appeared more interested in than men.

Top ten medical condition pages

Table 10 lists the top ten most popular medical pages. The top ten pages viewed accounted for 30% of pages used. Back pain proved to be the most visited topic, accounting for nearly 18% of the top ten pages viewed. Abnormal health rhythm, asthma, and migraine were the other medical conditions that accounted for more than 10% of the top ten pages viewed.

Table 8. *Top ten 'healthy living pages' viewed by place**

	Place				
	Truro	Oxford	Penzance	Stockport	Total
Alcohol	15.0%	22.9%	23.1%	18.8%	19.9%
Good eating	19.7%	19.1%	18.3%	20.4%	19.3%
Exercise	15.3%	14.8%	17.3%	20.6%	16.6%
Weight	7.1%	8.6%	11.0%	4.0%	7.9%
Cancer prevention	7.2%	5.6%	8.7%	10.2%	7.7%
Back pain	4.2%	8.7%	7.3%	7.1%	6.8%
Abnormal heart rhythm	6.8%	3.8%	5.9%	9.6%	6.2%
Smoking	12.6%	3.4%	3.6%	3.7%	6.2%
Stress	3.3%	9.4%	3.1%	2.9%	4.9%
Asthma	8.8%	3.6%	1.7%	2.6%	4.4%
Total	100%	100%	100%	100%	100%

*Excluding travel clinic pages.

Table 9. *Top ten healthy living pages viewed by gender*

	Gender		
	Female	Male	Total
Alcohol	14.7%	26.0%	19.9%
Good eating	20.4%	18.0%	19.3%
Exercise	16.2%	17.2%	16.6%
Weight	9.9%	5.6%	7.9%
Cancer prevention	8.3%	7.0%	7.7%
Back pain	5.9%	7.8%	6.8%
Abnormal heart rhythm	5.1%	7.6%	6.2%
Smoking	8.2%	3.8%	6.2%
Stress	6.1%	3.5%	4.9%
Asthma	5.2%	3.5%	4.4%
Total	100%	100%	100%

The key differences between kiosks were in regard to:

- back pain – 23.2% of all top ten medical condition pages used in Penzance and Oxford, but only 10.2% in Truro;
- abnormal heart rhythm – 24.3% of all top ten medical condition pages used in Stockport, but only 10.2% in Penzance,
- asthma – 21.3% of all top ten medical condition pages used in Truro, but fewer than 10% in other locations;
- enuresis – 12.6% of all top ten medical condition pages used in Stockport, but just 10.4% in Oxford, 7.1% in Truro and 5.6% in Penzance;
- iron deficiency – 10.3% of all top ten medical condition pages used in Truro and 3.6% in Oxford.

The key gender differences were (Table 11):

- 20.7% of males looked at abnormal heart rhythm pages as a percentage of the top ten pages viewed compared to 13% of females;
- 16.2% of females, as a percentage of the top ten pages viewed, looked at migraine pages compared to 3.6% of males;
- 14.8% of males, as a percentage of the top ten pages viewed, looked at brain tumour pages compared to 3.6% of females.

Clearly, comparisons of these data with national statistics on gender and migraine, heart rhythms and brain tumours, will form part of the analysis as the project progresses.

Table 10. *Top ten medical condition pages viewed by place*

	Place				
	Truro	Penzance	Oxford	Stockport	Total
Back pain – strain	10.2%	23.2%	22.3%	18.0%	17.9%
Abnormal heart rhythm	16.3%	10.2%	17.9%	24.3%	16.5%
Asthma	21.3%	9.5%	5.2%	6.6%	11.7%
Migraine	14.2%	7.2%	15.7%	3.8%	10.6%
Brain tumours	10.0%	12.1%	6.0%	4.1%	8.6%
Enuresis	7.1%	5.6%	10.4%	12.6%	8.5%
Iron deficiency anaemia	10.3%	7.6%	3.6%	7.3%	7.5%
Asthma in childhood	3.6%	9.5%	7.7%	6.6%	6.7%
Diabetic diet	2.1%	9.1%	6.0%	9.5%	6.3%
Carpal tunnel syndrome	5.0%	5.9%	5.2%	7.3%	5.7%
Total	100%	100%	100%	100%	100%

Table 11. *Top ten medical condition pages viewed by gender*

	Gender		
	Female	Male	Total
Back pain – strain	5.1%	21.4%	17.9%
Abnormal heart rhythm	13.0%	20.7%	16.5%
Asthma	13.4%	9.7%	11.7%
Migraine	16.2%	3.6%	10.6%
Brain tumours	3.6%	14.8%	8.6%
Enuresis	9.8%	6.9%	8.5%
Iron deficiency anaemia	9.6%	5.0%	7.5%
Asthma in childhood	6.5%	7.0%	6.7%
Diabetic diet	9%	6.9%	6.3%
Carpal tunnel	7.1%	4.0%	5.7%
Total	100%	100%	100%

Examining data from the perspective of the age of the user provides fascinating consumer insights (Table 12). The interest in the Cape Verde islands shared by the under 15s and over 75s is inexplicable. One explanation, at least for the former group, may be the animation which accompanies the travel pages, which are generally popular with the under 15s, although this still cannot account specifically for the attraction of the Cape Verde islands. The 16–35s and the over 75s share an interest in alcohol. Other features of the table are the 36–55's concern with back pain and the lack of interest shown by the over 75s in exercise and weight.

Conclusions

From the published research, it appears that a touch-screen or similar computer mediated GUI system can be beneficial for the conveyance of health information. Graham [11] and Buxton *et al* [9] both showed that user satisfaction with such systems is high, and Graham [11] and Patyk [17] also demonstrated, in different ways, that the information conveyed can increase user understanding. In the medical field, this helped towards a lowering of anxiety. Jones *et al* [15] indicated that the problem of 'information overload' could be combated with computerised information displays, because unretrieved pages are hidden, and the user is not faced with a mountain of paper to wade through. Buxton *et al* [9] and Velikova *et al* [16] showed the ease with which the systems could be used, and the latter also suggested that information input from patients could be as reliable as on

Table 12. *Top ten pages by age (all pages)*

Top ten pages viewed by age group					
	Under 15	**16 to 35**	**36 to 55**	**56 to 75**	**Over 75**
1	Good eating	Alcohol	Alcohol	Good eating	Alcohol
2	Exercise	Good eating	Good eating	Abnormal heart rhythm	Good eating
3	Alcohol	Exercise	Exercise	Exercise	Brain tumours
4	Brazil	Cancer prevention	Back pain – strain	Alcohol	Cancer prevention
5	Asthma in childhood	Weight	Migraine	Stress	Balanitis
6	Weight	Smoking	Weight	Cancer prevention	Colostomy
7	Cape Verde	Broken ankle – open	Smoking	Hip replacement	Smoking
8	Cancer prevention	Abnormal heart rhythm	Abnormal heart rhythm	Back pain – strain	Cataract lens replacement
9	Enuresis	Iron deficiency anaemia	Asthma	Knee replacement	Fiji
10	Egypt	Brain tumours	Cancer prevention	Weight	Cape Verde

paper. Jones *et al* [15] calculated that despite the apparent high costs of information technology as compared to hardcopy, multiple simultaneous access and lack of printing costs may make touch-screens more economical than producing and distributing leaflets, if one considers the longer term.

The early pilot data from the DoH funded study appears to confirm the popularity of kiosks with the general public, although it does show that there is a lot of variation in their use. Plainly age, gender, season, nature of site and its geographical location are all significant factors and their influence will be followed up in future work when the project moves to the investigation of all 70 plus kiosks. Platform – kiosk, web etc, is also likely to be influential and platform comparisons will prove to be very interesting here also.

References

1. Jones, R., McLachlan, K. and Bell, G. HEALTHPOINT: a public-access health information system. In: DeGlanville, H. and Roberts, J., eds. *Conference proceedings: current perspectives in health computing 1990.* Weybridge: BJHC Books, 1990.

2. Jones, R., Edgerton, E., Baxter, I. et al Where should a public access health information system be sited? *Interacting With Computers,* 5, 1993, 413–421.

3. Jones, R., Navin, L.M. and Murray, K.J. Use of a community based touch-screen public-access health information system. *Health Bulletin,* 51, 1993, 34-42.

4. Naven. L., Jones R., Kohli, H. and Crawford, J. How should we evaluate a public-access health information system? In: Richards, B., ed. *Healthcare Computing 96.* Weybridge: BJHC Books, 1996, 557–562.

5. Jones, R., Labajo, R., Soler Lopez, M., Sanz, J.J., Alonso, P. and Claveria, L.E. Evaluation of a Scottish touch-screen health information system in rural Spain. In: Bryant, J., ed. *Healthcare Computing 2000.* Weybridge: BJHC Books, 2000 (in press).

6. Pearson, J., Jones, R., Cawsey, A., McGregor, S., Barrett, A., Gilmour, H., Atkinson, J. and McEwen J. The accessibility of information systems for patients: use of touch-screen information systems by 345 patients with cancer in

Scotland. *American Medical Informatics Association Annual Symposium 1999: Session 66 – Consumer Health Informatics II.* 1999. www.amia.org/pubs/symposia/D005289.htm

7. Al-Barwani, F. *The development and evaluation of a patient workstation.* Unpublished PhD thesis. Glasgow: Glasgow University, 1998.

8. Jones, R., Pearson, J., Cawsey, A. and Barrett, A. Information for patients with cancer: does personalisation make a difference? Pilot study results and randomised trial in progress *Journal of the American Medical Informatics Association Symposium supplement.* 1996, 423–427.

9. Buxton, J., White, M. and Osoba, D. Patients' experiences using a computerised program with a touch-sensitive video monitor for the assessment of health-related quality of life. *Quality of Life Research,* 7(6), 1999, 513–519.

10. Douglas, G. Jones, R. and Naven, L. Schoolchildren's use of, and ideas for, a computer based health information system. In: Richards, B., ed. *Healthcare Computing 1995: conference proceedings.* Weybridge. BJHC Books, 1995, 139-144.

11. Graham, W., Smith, P., Kamal A., Fitzmaurice, A., Smith, N. and Hamilton, N. Randomised controlled trial comparing effectiveness of touch-screen system with leaflet for providing women with information on prenatal tests. *British Medical Journal,* 320, 2000, 155-160

12. Williams, E., Parker, A., Stoddard, J., Bomken, S. and Prabhu, V. Electronic letter: assessment of informed decision making. *British Medical Journal.* 2000. www.bmj.com/cgi/eletters/320/7228/155

13. Jones, R. and Craig, N. Letter: patient information systems are not more expensive than leaflets. *British Medical Journal,* 320, 2000a, 1212.

14. Jones, R. and Craig, N. Electronic letter: economic analysis of patient information. *British Medical Journal.* 2000b. www.bmj.com/cgi/eletters/320/7228/155 (visited 10 May 2000)

15. Jones, R., Pearson, J., McGregor, S., Cawsey, A.J., Barrett, A., Craig, N., Atkinson, J.M., Gilmour, W.H. and McEwen, J. Randomised trial of personal computer based information for cancer patients. *British Medical Journal,* 319, 1999, 1241–1247.

16. Velikova, G., Wright, E.P., Smith, A.B., Cull, A., Gould, A., Forman, D., Perren, T., Stead, M., Brown, J. and Selby, P.J. Automated collection of quality of life data: a comparison of paper and computer touch-screen questionnaires. *Journal of Clinical Oncology,* 17(3), 1999, 998–1007.

17 Patyk, M., Gaynor, S., Kelly, J. and Ott, V. Touch-screen computerised education for patients with brain damage. *Rehabilitation Nursing,* 23(2), 1998, 84-87.

18. Elting, L., Martin, C.G., Cantor, S B. and Rubenstein, E.B. Influence of data display on physician investigators' decisions to stop trials. prospective trial with repeated measures. *British Medical Journal,* 317, 1998, 1527–1531.

Screen-based information design: Eforms

Alistair Ross
Department of Information Science, City University, London EC1 OHB
ross_alistair@hotmail.com

The study surveys and delineates the processes involved in screen-based information design. This is specifically in relation to the creation of electronic forms and from this offers a guide to their production. The study also examines the design and technological issues associated with the transfer, or translation, of the printed form to the computer screen. How an Eform might be made more visually engaging without detracting from the information relevant to the form's navigation and completion. Also, the interaction between technology and (document) structure where technology can eliminate or reduce traditional structural problems through the application of non-linear strategies. It reviews the potential solutions of incorporating improved functionality through interactivity.

Introduction

If we disregard the hyperbole regarding a digital Nirvana or Armageddon, just what do we really have when we boil it down to the bare bones? We have a global network of computer networks transferring data around the world in the blink of an eye. Computer manufacturers are no longer simply in the business of making computers they are now in the business of processing information. Internet Service Providers do not just provide access to the Internet, but content along with an increasing array of services such as email, games, shopping and search engines.

'The new electronic interdependence recreates the world in the image of a global village.' [1, p. 67] With the advent and development of the Internet and the World Wide Web McLuhan's prophecy for a global village became a reality, one that we must learn to navigate, participate in and understand. Concurrent with the growth of this new virtual domain there has been a necessary and corresponding development in the range of digital documents. Or has there? The focus of interest here is the translation of documents from the material (paper) to that of the virtual (digital) medium.

Cognitive psychologists, human-computer interaction (HCI) engineers, information architects, programmers, and graphic designers, all do their bit to make websites engaging and appropriate to the nature of the content whilst working within the limitations of a new medium and related technologies.

Obviously, it would be disingenuous to suggest that the above description is anything but a gross simplification and does not reflect the effort, diligence and dedication of the exponents of each discipline. Also, there are many areas of investigation and disciplines not mentioned. There is diversity of effort targeted at this area, but with little or no constructive, or more accurately, visible success as far as the general public is concerned. Naturally, there are exceptions, but they are just that. Exceptions!

Scope and definition

On initial consideration there are three distinct spheres of investigation in connection with the larger scheme – those of technology, structure, and process. Information technology, understandably, is located within the 'technology' sphere, and held within the 'structure' sphere is information architecture, human-computer interaction [2] (HCI), and graphic design. In this instance graphic design may be considered a subset of HCI.

> HCI: the study of how humans use computers and how to design computer systems which are easy, quick and productive for humans to use [2, p. 12].

However, whilst these fields are discrete there is overlap where appearance and intent coincide to determine the measure of usability and where any defining line blurs. Indeed, technology and structure are encompassed within process. It would be reasonable to define a form as an interrogative document. A document, that by its function is a receptacle for the data received.

The proliferation of the Web has resulted in an explosion in the amount of information (good, bad and indifferent) available both in print and on the Internet. Countless books and web-guides have been produced along with the inevitable instructional courses for the design of web pages and sites, various software, HTML coding, and programming, etc. Each of them dealing with the various components of the process, each giving differing levels of content and technical information, targeted at different audiences and overall generally leading to chaos and confusion.

Evaluation

The digital document can be evaluated on a number of levels and contexts, from the aesthetic presentation of the graphic constituents (text and images), to the ease of site navigation and the accessibility of information – the architecture of the website. Understandably, there are a variety of ways in which to evaluate human-computer interaction systems and the software constructs running on them. The evaluation process was directed with two main objectives in mind: to determine the effectiveness of an interface in use, and to provide a means for suggesting improvements.

Heuristic evaluation

Sadly, the usual evaluation methods were unsuitable. One evaluation method that did suggest itself – heuristics – was only partially appropriate.

> Heuristic evaluation as described here, however, is a systematic inspection of a user interface design for usability. The goal of heuristic evaluation is to find the usability problems in a user interface design so that they can be attended to as part of an iterative design process. ...Heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles [the 'heuristics'] [3, p. 155].

A typical set of heuristics would be the usability principles proposed by Jakob Nielson in *Usability Engineering*. 'Heuristic evaluation is done by looking at an interface and trying to come up with an opinion about what is good and bad about the interface' [3, p. 155]. Nielson's 'usability principles' provided a basic framework for the evaluation.

Nielson's usability principles

1. **Visibility of system status**
   The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. **Match between system and the real world**
   The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3. **User control and freedom**
   Users often choose system functions by mistake and will need a clearly marked 'emergency exit' to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4. **Consistency and standards**
   Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. **Error prevention**
   Even better than good error messages is a careful design that prevents a problem from occurring in the first place.
6. **Recognition rather than recall**
   Make objects, actions and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the sys-

tem should be visible or easily retrievable whenever appropriate.

7. **Flexibility and efficiency of use**
   Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
8. **Aesthetic and minimalist design**
   Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9. **Help users recognize, diagnose and recover from errors**
   Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
10. **Help and documentation**
    Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Nielson's principles did not completely satisfy the evaluation need as their focus was usability. Consequently, it was necessary to seek a suitable adjunct to complement them.

David Siegel [4] discusses the development of websites in terms of 'generations'. This set of criteria enabled the combination of design aesthetics with Nielson's usability principles to create the necessary inclusive evaluation criteria. Siegel's generations provide a means of evaluating the evolution of websites, he contends that we are currently at the third generation of website evolution.

First-generation sites are linear and purely functional. Very basic in appearance with no frills and little, if any, aesthetic consideration. The page displays top-to-bottom, left-to-right sequence of text and images interspersed with basic data dividers such as bullets and horizontal rules. With edge-to-edge text setting that runs for several pages. These sites were designed by technical people usually with other technical people in mind and are still produced.

Second-generation sites are the result of a set of extensions to HTML. Basically, these sites are first-generation sites with icons replacing words, tiled images replacing the grey background, buttons with bevelled edges, and banners replacing headlines. They use a top-down, bullet-list and menu-driven model to present a hierarchy of information. 'A third-generation site is wrought by *design*, not technological competence'. [4, p. 15]

Third-generation sites pull visitors within by using metaphor and well-known models of consumer psychology. The cleverness of third-generation designers is not technical but visual. Design is the difference. The designers expend time making their pages enticing to their audience.

Document structure

Structural constraints

There are structural constraints inherent in all documents and specifically in forms – digital and print-based. These constraints are placed on the document by its nature, making the traditional form's structure invariably linear. You are forced to progress through the form answering the questions in sequence. You are not expected, or required, to answer questions out of sequence or in any other manner other than the prescribed one. For example, if it becomes necessary to refer to guidance notes in order to answer a question appropriately (e.g. passport application form) the user must leave the form and read the notes for clarification. The process becomes fragmented and is no longer linear.

Here we arrive at the major difference between the two types of form. The printed form is essentially linear in structure, but the Eform while it shares many of its precursor's characteristics has no real need for such a structure. Indeed technology makes it possible for it to create its own structure, hence the importance of interactivity as a defining factor. In theory, it is possible to create a document that will allow the user to move through it in whatever manner they choose.

The form and the type of data sought determine whether this is practical in terms of data collection.

In the printed form the user is often only required to answer specific questions that apply to his situation and not others. He will be instructed to ignore certain questions and go on to the next relevant question, made relevant by his previous answers. This technique is commonplace and is known as *routing*. The user still sees the non-relevant questions, as they are printed and cannot be removed, leaving the potential for error. With the Eform the irrelevant questions are removed (more accurately, not shown). Automatically progressing to the next relevant question using hyperlinks. Therefore the potential for error is reduced, or eliminated – you cannot fill in the wrong question if it is not there for you to fill in!

Technology and limitations

The printed form is inert – whether it is filled in or not; the form merely sits there waiting until you decide what to do with it – old technology. In its defence it must be said that the printed form does have many advantages. It does not need expensive electronic hardware and software to view or complete it – a biro will suffice. However, where the printed form ends is where the digital form begins! Interactivity is the key word; once the Eform has been created it has the potential to interact with the user.

Eforms can incorporate the necessary notes and information to answer questions. Access to the guidance notes can be triggered in several ways. For example, should the respondent exceed a pre-specified time limit in answering a question, he could be asked if help is needed, or tick a help box located within the question area. So instead of two static (print) documents, the form and the notes, there is one dynamic (digital) document.

Translation issues

Both media require a substantial level of technology in order to create a document. Although, it is debatable which requires the greater allocation of resources as both require extensive amounts to be committed. The printed form requires a computer and the software to produce the design, followed by necessary reprographic equipment to print it and then a distribution system. Its screen-based counterpart requires a computer and the software to enable the programming, and a server to house the document and receive the respondent's data. This simplified outline ignores the efforts of the designer and programmer to engender the whole thing.

The essential difference between the printed and the screen-based forms appears to be in the nature of their use. The printed form has the advantage of being ready for completion once the document is delivered. The user can complete the form whenever and wherever they wish. On the other hand, the screen-based form is exactly that 'screen-based' it can only be completed if there is a screen (and computer) there to allow the process to take place.

Conversely, the screen-based version has the advantage over print when it comes to distribution and return. The printed form needs to be physically transported to and from the user. The Eform is transferred virtually instantaneously (via the web) to any location having the appropriate technology.

A second, and extremely compelling, advantage that Eforms have over their print counterpart is that the data received from the Eform can be an element of an integrated data collection and processing system.

The survey

An initial search for digital forms was conducted on the Internet and a variety of Eforms located. The survey revealed that there are four basic types of digital form. Each category defined by the software technology required to create the form:

- The word processor file
- The Portable Document Format (PDF) file.
- The Web (HTML and/or other scripting program).
- Stand-alone software.

The word processor file

The word processor file is perhaps the simplest of all the categories because the file is often nothing more than a Microsoft Word (or similar) file intended to be printed out and then completed, rather than be completed on screen. This file has the advantage of being cheap and simple, requiring nothing more than the word processing software package to produce and view it. It can be attached to email and sent to the respondent without difficulty. At this level there is no interactivity.

There are several flavours within this category, each indicating a slight raising of complexity from the basic Word file. First there is the basic Word file that replicates the printed version on screen; in theory, it is possible to complete this form on screen but not advisable as entering data changes the initial layout (with every keystroke). The second is where the file uses the 'tables' formatting to create the form. Here it is possible to complete the form without causing changes to the layout. However, it must be noted that in both of the first two variants the file may be altered on the respondent's computer should the original font not be available. In the first example there is a limited possibility of interactivity, but in the second it is possible to interact with the form (complete it).

The final variant is that of the imported image file. The reasons for this are that the image may contain some type of formatting not possible in Word, or the document layout is required to remain unchanged on the respondent's system. Also, the original document may have been created in software not widely available and it was realised that a Word file option was sensible. It is not possible to complete the form; it is necessary to print out the form to complete it. Again, nil interactivity.

The Portable Document Format (PDF) file

The Portable Document Format (PDF) [5, 6] file is a variation on the previous file type in that it creates a form document that cannot be altered by the respondent's computer. Thus it resolves the problem of how to distribute and receive a uniform document. However, there is an obstacle in accessing documents produced using this format. The PDF format is a proprietary format. To read the file Acrobat Reader® [7] software must be downloaded (free) from Adobe's website. Creating a file requires the author to buy the Adobe Acrobat® software. PDF is a double-edged sword: the ability to create inviolate documents also means that it is not possible to fill in the form. The document must be printed, and then completed, a retrograde step. The document may also be attached to email. There is nil interactivity.

The Web (HTML and/or other scripting program)

The Web (HTML, etc.) file using 'radio' buttons: this example is a true Eform in all senses – created, read and replied to electronically. Whilst it may be more impressive in terms of technological delivery this file type does tend towards the visually mundane. The radio button is a simple visual device used to indicate that a choice has been made. The choice can be a simple yes or no, or can be a scale where a preference can be indicated. Selection can be restricted, so that only a yes or no may be ticked. Once the selection has been made the 'submit' button is pressed and the data is sent to the appropriate location. This type of form is common on the Internet. Interactivity consists of completing the form and, if the document has error messages, to prevent incorrect selection. In many instances there are no such safeguards put in place.

The Web (HTML, etc.) file using text fields. This example is also a true Eform. The text field is a simple visual device using the metaphor of empty text boxes with textual data entered via the keyboard. Again, the document may have error messages to prevent the incorrect, or inappropriate, entry of text, where text length is restricted or only numbers may be entered, or that no value has been entered at all. Again, frequently there are no safeguards instituted to prevent incorrect completion of the form.

Stand-alone software

The stand-alone software package is inevitably the most complex of all the category types. This type of package is usually custom made to suit a particular purpose in data collection. It is also used to ensure that only specific types of data are gathered. The data amassed are usually interrelated or cumulative in nature. The package is provided either on disc or can be downloaded It may contain several modes of entering the data: point and click, radio buttons, text fields, etc., depending on what type of data is required at any particular point. The advantage of such packages is the ability to create 'audited' documents that exactly adhere to the predetermined requirements of the client's query. In terms of interactivity this type of document is the most likely to exhibit true interactivity. The user will be guided on how to complete the form, not just with error messages but with a range of measures. The form reacts to the user's actions or lack of action. Obviously, there is likely to be a sliding scale of interactivity depending on the software.

The EVR (Electronic Version Return) [8] tax return form produced by the Inland Revenue is the best example of an Eform encountered during the study. Granted the aesthetic is hardly the most engaging, but context must be weighed against this – will the taxpayer be happier doing an unpleasant task if it is made more jolly, or will it create animosity because money was spent making it so. The interactivity was well thought out and virtually all potential errors and problems were avoided before they happened. One suspects the main reason for creating this software was that the Inland Revenue needed to update the means of providing a way for people to enter their tax information in this modern age. The main problem being the necessity of ensuring uniformity. The simplest solution was for the Inland Revenue to provide the software, thus avoiding the nightmare of commercial software vendors providing partial or incorrect solutions.

Technology

Technology within the context of this study comes in two shapes: hardware and software. Hardware is the physical bits and pieces that can be seen and touched: the computer, the screen, the mouse and keyboard. Also it refers to the telecommunications systems that exist elsewhere to enable the entity known as the Internet to exist. Software is the programming in the hardware that tells the hardware what to do. These explanations are oversimplifications. but necessary if we wish to progress without getting bogged-down in minutiae.

The process

The basic premise of a web page or document is that it is created and then exists on a server. This is referred to as a 'client-server' system: a networked system where the processes are split between server computers, holding files for distribution, and client computers that request files and services from the servers [9].

So the web document is created on a (client's) computer and is then transferred to a server computer on the network (e.g. the Internet). Anyone (matching set criteria, if on a private network) can then access the document.

To construct the document it is necessary to create it using some form of programming language that describes the page and its contents (e.g. HTML). To make the relevant parts of the page interactive it is necessary to program the interactivity using a scripting language (e.g. DHTML, Perl, JavaScript, etc.) [9].

Hardware

Resolution

This is a physical limitation placed on the screen by technology, referring to the sharpness and clarity of an image. The current standard screen resolution is around 72 dots per inch (dpi). When compared to that of print the screen resolution is ridiculously crude. Print is around 2,000 lines per inch.

A question of colour

The screen of a monitor is comprised of pixels (picture elements); each pixel consists of three coloured phosphor dots – red, green and blue (RGB). The RGB colour system is additive, the colours are combined optically producing the full range of colours [10].

The browsers each have their own particular idea of what they consider an acceptable colour. For example Microsoft's Explorer cannot render some colours, bringing the palette down to 212 colours from 256. So for obvious reasons, it is best to use a 'browser-safe' colour palette to ensure that the colours used are the ones viewed by visitors. Where colours used cause problems, the offending colour is rendered by 'dithering', creating an unpleasant speckled effect [11].

Software
In this instance we consider software where it has been used to facilitate construction or use of the form. The key components we need to consider are those of hypertext links, programming languages and the systems that use them.

Proprietary software
This software's sole purpose is to create forms and questionnaires from scratch, or from an original printed form. The software can automate data entry routines to make filling in Eforms more efficient through routing and automatic connection to databases [12, 13, 14]. Such software provides a ready-made but at a price! It is not cheap and may interact adversely with other parts of the system.

Hypertext
Hypertext is important because it is the base technology upon which the functionality of the Internet rests. Hyperlinks enable the user to go from one website to another, or from one part of a document to another. This movement can be intentional, triggered by the user, or can be made unobtrusive in that the transfer may be made without the user's knowledge.

Programming
There are a number of programming languages and scripts relevant to the Internet. Perhaps the most commonly known is that of Hypertext Markup Language (HTML). HTML was set up as an attempt to provide a universally accepted, cross-platform standard language for displaying information, text and visuals on the web [15]. HTML is a set of codes used to describe the document's structure and allow hyperlinks to other documents on the Internet.

There is nothing dynamic or interactive about HTML documents – they are termed 'flat' files [9]. They describe the structure of what is on the screen and little else. This is where DHTML (Dynamic HTML), amongst others comes to the fore; any object on the page can be made to react to user actions, cause sounds to be triggered or the format of the page to be modified, etc. DHTML interacts with programming languages and scripts to allow these actions.

The other half of creating interactivity on the web page is the necessity of adding program or script instructions. If you want your site to do something interactive it is necessary to run a separate program outside of the web server. This is where CGI (Common Gateway Interface) [9] comes in – it is the interface between these applications and the web server. These external programs can handle incoming forms, retrieve records from a database, save data to a file, and so on. As long as you conform to the CGI specification, there is a wide choice of tools for writing CGI applications. Typically, a scripting language is used, although you can write CGI applications in just about any programming language [16].

Guide: Screen Design

There is no ultimate design solution when it comes to screen design or any other type of design. Consequently, this article focuses primarily on what to avoid rather than what to do. Remember design is a process, it is an art not a science, it is possible to follow the guidelines and still produce something that is best consigned to the waste bin.

Typography
Most file formats do not contain the font information within the file – rather they include information about what fonts are used in the document. In order for those fonts to be displayed they must be installed in the system of the computer displaying the file. There is no such thing as 'web safe' or 'browser safe' fonts [11].

One solution to this problem is to create type as graphics, to make type an image. The browsers will simply treat it as another graphics file and it will remain unaltered. The downside to this is that graphics files take longer to download than ordinary text files. An alternative to this is to use downloadable fonts. The latest browsers have incorporated a feature to allow this. This is still in its infancy, as different browsers require a different font format [9].

Legibility and readability

Type is adversely affected by the relatively low screen resolution. So it is unwise to have type smaller than 11 points in size or italicised type because the screen resolution makes it difficult to read. Some fonts are more legible than others; select those for use by comparing them onscreen. Alternatively, use anti-aliased type. Anti-aliased type is a technique used to visually 'smooth' the shapes in type by inserting pixels of intermediate colours along boundary edges between colours. It removes the jagged edges of larger type characters. At normal viewing distances anti-aliasing gives the impression that the type is rendered at a higher resolution. Image editing programs such as Photoshop create anti-aliased type [17], but as a graphic rather than plain type. Lower case type works best in most circumstances and is the easiest to read. Careful use of emphasis adds structure to the text indicating what is or is not important.

Text alignment is another important component in type readability. As there is unlikely to be a large amount of text in any onscreen form it is unnecessary to detail the failings of various styles of alignment. Whatever style of alignment is used make sure that it is used consistently and that it is appropriate. One other factor in text setting is that of line length. Line length is a constant source of irritation – many people view web pages using the 'full-screen' browser option. The line length of the text fills the screen making it virtually impossible to read it without difficulty. It is possible to limit the line length onscreen using HTML.

Finally, it must be remembered that on top of the low screen resolution problems there are also the added problems of screen flicker, image and colour distortion, and light reflection.

Colour

Colour can work in two ways. It can attract or repel a visitor's interest, so use it carefully. Contrasting colours can and should be used, but sensibly. There is a world of difference between colours that produce a helpful contrast to aid easier reading and clashing colours that only hinder communication.

With Eforms there is no need to stay with the neutral or similar safe colours, or the usual pale grey background. However, remember to ensure that the method of data entry (text box, etc.) is sufficiently 'visual' enough to survive in a colourful background.

Background colours should not be too bright or intricate as they will swamp everything else, given that it is usually the largest single area of colour. Nothing should distract attention from the text or other important information, or cause eyestrain and irritation, resulting in difficulty of viewing for the visitor. Again, it is best to use a 'browser-safe' colour palette to ensure that the colours used are the ones viewed by visitors. If this factor is ignored then an offending colour can create unwanted effects to images and backgrounds [18].

Graphics and Images

Images or graphics files are another vital component of the web page, although images do not usually feature in forms. Graphics often do feature in that the site wants the user to know whose site it is (by using a logo). Remember that graphics or image files must be kept as small as possible due to the amount of time they take to download.

Sound

Sound is another element that can assist in the interactivity of the form, for example to sound an error warning whenever an incorrect action has been made. This is best used in conjunction with a dialogue box explaining the error and what actions are necessary to recover from the error. Remember that constantly sounding beeps can be a distraction and therefore use only with caution.

Layout

Layout is more than merely the location of items on a page and the space in between and around those items. It is a synergy of the page content that produces the page. How well or badly the elements are integrated determines the page's visual viability. The key thing with layout is to keep it simple, avoid overly complex forms. If it is necessary to have multiple sections then use multiple pages rather than trying to cram everything on a single page. Use a grid system [19] to place the components, creating a cohesive unit from a collection of disparate parts.

Create structure by using differing type headings (creating levels of emphasis) to indicate hierarchies along with sufficient space to separate the blocks of text, text fields and images. Use type, headings, colour, button location and space in a consistent manner [20].

Vitally important to forms and questionnaires are text fields that are large enough for text to be entered and have the text remain legible. Several major commercial sites fail in this. If the visitor cannot read what he is writing how is he to avoid errors, or make corrections if an error is made? This is a particular failing of sites with their own search-engines, a tiny text field showing illegible type! Alignment of the text fields should be consistent and fit within a systematic framework. Ignore this at your peril, alienated users do not stay long enough to complete the form and so the only feedback you will receive is silence.

As Waters states, '...don't forget to explain things to your visitors along the way: what you want from your visitors and what they get from you' [21, p. 115]. This is especially true if they want to get the appropriate response to their questions.

Context

Context is everything, especially within the visual world. If a particular visual style or format is produced for one company's website it does not automatically follow that the design will be suitable for another – what is good for Disney will not suit the Pentagon.

Also, it would be considered out of context (or simply offensive) if a questionnaire enquired about the sexual proclivities of the respondent, or what they had for breakfast, given the sole intention was to create a database on the number of cars owned by each household.

Context is the reason why it is so difficult to produce definitive answers to how to design something. It would be fine to be able to say that you must do this, or not do that, to produce the ultimate Eform, but it is not possible. Alter the circumstances, or any of the factors involved, and you alter the context.

Iterative testing

Iterative testing is something that applies to all of the above – ignore it or fail to include it at your peril. Nielson [3] makes the case for iterative testing and design of interfaces in *Usability Engineering*. He suggests that testing should form a major part of the design process. Siegel [22] (as does Waters [21] and Weinman [18]) also makes the point that designs should be tried and tested before being released upon an unsuspecting world.

Kristof and Satran [23, p. 119] state that, 'Audience research becomes user testing when you can stop asking people what they think about an idea and let them try the real thing.' They also make the point that testing is also used to determine which design is to be used in preference to another rather than merely to avoid or remove defects.

If people are to view the site and complete the questionnaire then it is vital that nothing interferes with the process, all barriers to this end must be removed, before the form is distributed. Remember that no amount of stunning aesthetic design and wonderful information structure will counteract bad communication or get people to answer the questions you did not ask!

Conclusions

Structure and analysis

As originally stated, there are three distinct spheres of investigation in connection with this study: those of technology, structure and process. Let us consider the three spheres and how they impact upon the Eform.

The technology sphere features large, without it there would not be any Eforms. The technology involved in the creation of an

Eform has already been outlined and how it is used to facilitate improved interaction with respondents. The two technologies (print and telecommunications) present different advantages to the client. The printed form has the advantage of being ready for completion once the document is delivered. On the other hand, the Eform is exactly that 'electronic' it must be completed where there are suitable facilities. However, the Eform has the advantage over the print when it comes to distribution and return delivery. Printed forms must be physically transported to and from the user. The Eform can be transmitted instantaneously (via the Web) to any location linked with appropriate technology. Another advantage of Eforms over their print counterpart is that the data received can be preformatted for transmission to databases, or can be a component within a data collection system

Structural divergence between the two forms is in the nature of their use. The immediately striking difference is that the Eform has the capacity for interaction whereas the printed form does not – this is a significant advantage. Both types of form are interrogative in nature, but the Eform is able to act or react if certain criteria are met (or not met)

This means the inclusion of notes to questions, help dialogue boxes, or (invisible) routing, etc., creating a different document structure. The printed form is basically a linear document. The Eform retains this linearity, but it has the potential to be completed in whatever manner suits the user best – software permitting. Here technology contributes in a positive manner to the whole process rather than using it because we can.

Process, the final sphere, encompasses technology and structure. As previously indicated, the technology associated with Eforms has effectively increased the speed of operation involved in the collection and processing of data by removing the physical barriers that exist with print. The creation of Eforms that are imbedded within the data-processing system rather than being an adjunct to it further removing those barriers.

The ability to create hypertext documents over the old linear forms, the capacity to use 'invisible' routing (invisible to the user) that effectively creates customised forms for the respondent.

The forms that react if certain criteria are met (or not met), the integration of notes with the questions, or help dialogue boxes that appear after a set time if the question has not been answered, etc.

Basically, the process has been changed from what was a purely linear process to one that allows the restructuring of the document to suit the enquirers and the user. Along with the ability to accelerate the whole process and the capacity for interaction through the use of technology. However, much remains to be done in the way of improving the design: aesthetic and structural of the Eform.

Form design is too often left to the last moment, with little consideration of their workings or of the respondents – this approach is shortsighted! Design in whatever context is not a bolt-on to be applied after the fact. It is integral to the whole process!

Boldly going...

We are at the frontier. It is still early days in the development of the digital medium. Both the medium and its related technology still have to deliver on their promise. Much was promised for the radical improvement of communication, extending the dissemination of information. A brave new world! Some of it has been delivered, but, to paraphrase Frost, *we have miles to go and promises to keep*.

As television is the elder relative, it seems reasonable to use it as a comparison. Colour on television is now taken for granted, but it took a while to arrive. Colour televisions have been available (in the UK) since the 1960s, but it was not until the early 1970s that the first network colour broadcasts occurred. It took several years for colour TV to look normal, and websites rapidly caught up in only a few years. Colour use in newspapers is still in its infancy despite, or perhaps because of, colour being used in magazines for some considerable time.

Just as the development of desktop publishing spawned millions of bad (would be) designers so the popularity of web authoring programs has put the means of production into the hands of millions – millions of bad publishers.

What the future will have to offer us in the way of improvement is anyone's guess. The possibility of fully working speech-recognition is already with us, with improvements almost being made daily – the completion of Eforms without a keyboard or mouse?

However, before we rejoice in all that technology is about to offer us we should be aware of what is actually happening in the real world. There is at present an increasing gap between the technology 'haves' and the 'have-nots' – one that must be bridged if we are to realize the full potential of the digital medium. Until computer access is widespread and cheap then little will change except, perhaps, in the corporate world. The cost of being online is cause for heated argument between consumer groups and British Telecom – there is little point being linked to the Internet if you cannot afford the phone bill!

Bad design

'Bad design will always be with us...' [24]. It is perpetrated by both the designer and by the layman. Sadly, the purchase of the relevant software packages (design and/or web authoring) does not confer a sense of aesthetic, design training and experience, or the intelligence to know when something is best left to the experts! This is not to say that design is the province of only those with design training, we must remember that Harry Beck, a draughtsman with no design training, designed the London Underground map. A design classic! [25] It has yet to be explained that if forms are so simple how is it that there are so many appalling examples produced? Even in the case of the printed medium this is still the case. Like many things that appear easy to do, they are deceptive because the experts are very good at what they do – that is why they are experts. Unfortunately, it is just that there is more bad design than good. Inspired designs are not noticed because they are too busy solving the problems and making things happen without stress.

Information design

Emphasis has been placed on the need to produce documents that are well structured and easy to read and navigate. All these are a part of, or connected with, information architecture and design. The systematic approach to the design and layout of web pages, and the forms contained within, falls within the remit of the information designer. The imposition of structure upon the information is vital. If a visitor has difficulty in reading the form or finding their way around the site then the architect/designer has failed. 'Information designers talk about 'granularity' – putting just the right amount of information in front of the user at any one time' [4, p. 27]. As discussed earlier, information architecture, human-computer interaction (HCI), and graphic design are constituents of the 'structure' sphere. Inextricably interconnected to and overlapping with each other, all concerned with the structuring and presentation of data. Give consideration to the user and his environment [4].

Interactivity

Interactivity is the key ingredient to the successful production and use of Eforms and most other types of digital documents. It is not proposed that forms become some kind of kitsch entertainment, with no end of beeps, flashing icons, etc. Rather, that they require even greater thought and attention to detail than traditional forms. And that proper thought be given to their function and to the processes involved in their design, construction and use. Only use what actively assists or helps to achieve the set objective and only that, all else is superfluous or vanity.

Security

Finally, you have successfully designed and created a fully working form, one that people are happy to complete and you have useful data flowing onto your server. What protects all that data from the data pirates who want to steal it, or hackers wanting to demonstrate their power to disrupt it and your website? Some kind of security precautions should be taken to protect the data, en-route (encryption) and on arrival at the server (firewall).

Unfortunately, the various aspects of security are outside the scope of this study – this particular facet constitutes a research project all by itself!

Encryption: Where data is translated into a code unintelligible to anyone without the encryption key. Encryption is the most effective way to achieve data security.

Firewall: A system designed to prevent unauthorised access to or from a private network. Firewalls can be implemented in both hardware and software, or a combination of both [26].

References

1. McLuhan, M., Marshall, and Fiore, Q. *The Medium is the Massage*. London: Allen Lane, Penguin Press, 1967.

2. Preece, J., ed. *A guide to usability. human factors in computing*. Harlow: Open University & Addison-Wesley, 1993.

3. Nielson, J *Usability engineering*. London: Academic Press, 1993.

4. Siegel, D *Creating killer web sites*. 2nd edition Indianapolis: Hayden Books, 1997

5 Adobe Systems Incorporated. *Adobe PDF*. http://www.adobe.com/products/acrobat/adobepdf.html

6. McKinley, T *From paper to web: how to make information instantly accessible*. Indianapolis: Adobe Press, 1997.

7. Adobe Systems Incorporated. *Acrobat*. http://www.adobe.com/products/acrobat/main.html

8. Inland Revenue *Self Assessment*. 1999. http://www inlandrevenue.gov uk/sa/

9 DiNucci, D. et al. *The designer's guide to a new medium: elements of web design* 2nd edition. Berkeley, California: Peachpit Press, 1998.

10. Gotz, V *Color & type for the screen*. Switzerland: RotoVision, 1998

11. Slocombe, M. Colour Blind. *Internet Magazine*, 59, 1999, 101–2.

12. Caere Corporation. *OmniForm* http://www.caere.com/products/omniform/

13. JetForm Corporation. *Home Page*. 1999. http://www.jetform com

14. Q&D Software Development. *WebForm*. 1997. http://www.q-d.com

15. Cotton, B. and Oliver, R. *Understanding hypermedia. from multimedia to virtual reality*. London: Phaidon Press Ltd., 1993.

16. CNET.com *Builder.com. The ABC's of Common Gateway Interface*. 1999 http://www builder.com/Programming/CGI/ss01 html

17. Lynch. P J. and Horton, S. *Web style guide. basic design principles for creating web sites*. London: Yale University Press, 1999.

18 Weinman, L. *Designing web graphics 2: how to prepare images and media for the Web*. Indianapolis New Riders Publishing, 1996.

19. Garcia, M.R. *Redesigning print for the Web*. Indianapolis. Hayden Books, 1997.

20. Fleming, J *Web navigation: designing the user experience*. Sebastopol, California: O'Reilly & Associates, 1998.

21. Waters, C. *Universal web design*. Indianapolis: New Riders Publishing, 1997.

22. Siegel, D *Secrets of successful web sites. project management on the World Wide Web* Indianapolis: Hayden Books, 1997.

23. Kristof, R and Satran, A. *Interactivity by design: creating and communicating with new media*. Mountain View, California: Adobe Press, 1995.

24. Ennis, F. Web Design: the paper paradigm. *Monitor*. 1999. http://www.monitor.ca/monitor/issues/vol3iss9/feature2.html#2

25 Garland, K. *Mr Beck's underground map*. London: Capital Transport Publishing, 1994.

26. Internet.com Corporation. *Webopedia*. 1999. http://webopedia.internet.com

rather to discuss concepts: the difference being, according to Belkin [4, p. 58], a definition 'says what the phenomenon defined is, whereas a concept is a way of looking at or interpreting the phenomenon'.

In their recent paper, McCreadie and Rice [5] review concepts of information proposed over the last fifty years. A summary of the concepts they consider is given below.

- **Information as a representation of knowledge**
  Information is stored knowledge. Traditionally the storage medium has been books, but increasingly electronic media are becoming important.
- **Information as data in the environment**
  Information can be obtained from a range of environmental stimuli and phenomena; not all of which are intended to 'convey' a message, but which can be informative when appropriately interpreted
- **Information as part of the communication process**
  Meanings are in people rather than in words or data. Timing and social factors play a significant role in the processing and interpretation of information.
- **Information as a resource or commodity**
  Information is transmitted in a message from sender to receiver. The receiver interprets the message as intended by the sender. There may be added value as the information is disseminated or exchanged.

Information in context

The model presented below rests on the assumption that information cannot be evaluated without an awareness of the context in which it is being interpreted. This assumption leads to a model comprising three components.

1. Readership context

The context in which a message is received and interpreted. The reader is any system which derives (or attempts to derive) infor-[illegible]ion from a message. A system may be a [illegible] organism, a community, or

2. Authorial context

The context in which the message originates. The author is any system that transmits (intentionally or otherwise) a message from which a reader can derive information.

3. Message

The means by which information is transmitted. It may be written, spoken, facial expression, pheromonal, etc.

These components are described more fully below.

Information as data in the environment: reading the signs

Many information scientists accept that information is a property of all living organisms [6, 7]. It is not unreasonable, therefore, to illustrate the prime importance of context with an example from biology. It has long been recognised by zoologists that there is an association between brightly coloured markings on an animal and unpalatability or toxicity [8]. Predators learn to associate such markings with unpleasant experiences and so are less inclined to attack similarly marked creatures in future. So a potential predator, seeing the markings of a brightly coloured male butterfly, will derive information about the insect's suitability as a food source. A female butterfly of the same species will derive no such information. She might, however, find the markings useful in assessing the male's quality as a mate. Clearly, therefore, both the predator and the female butterfly derive information from the markings, but the message of the markings depends on the context in which it is read.

The importance of context in the interpretation of information has long been recognised. The above example, however, indicates that, unless it is assumed that butterflies and birds have knowledge, knowledge is not necessary for a signal to be informative.

Information as part of the communication process: the authorial and readership contexts

Messages exchanged between humans frequently differ from those discussed in the

A definition of information

A.D. Madden
JFS, Learning Resources Centre, 175 Camden Road, London NW1 9HD
admadden@hotmail.com

One difficulty faced by students on many information management courses is the lack of any attempt to teach concepts of information. Therefore, if a core module does not fit in with a student's existing concept of information, it can make it hard for the student to recognise the relevance of that module. This paper addresses that problem by summarising concepts of information, and by presenting a simple model that attempts to unite the various concepts listed. The model is based on the idea that the meaning in a message depends on the context in which the message originated (the authorial context), and the context in which it is interpreted (the readership context). Characteristics of authors, readers and messages are discussed. The impact of the 'knowledge' of 'information' users, and of their community, is considered. Implications of the model are discussed. A definition of information is suggested, which attempts to encapsulate the nature of information implied by the model.

Introduction

Following a return to full-time education after several years in scientific research, the author recently completed a MSc in Information Management. As student representative on the teaching committee the author was able to gain insights into some of the difficulties experienced by the course organisers. A disproportionate number of complaints were received about one module in particular. The module in question, though no more intellectually demanding than any other, was technically more demanding. It was a core module, so students had to take it, and many resented the fact. One common criticism was that the relationship of the module to the course as a whole was unclear. It became apparent that the issue of whether or not the module was appropriate was dependent on a student's concept of information. It was also clear that concepts of information were not being taught on the course. Discussion with students on courses elsewhere in the UK suggested that this is not unusual.

The paucity of theory in information science has often been commented on [1, 2], and has led to a pragmatic approach to both the teaching and practice of information science. The resulting piecemeal view of the subject has, not surprisingly, led to problems of the kind described above. The purpose of this short paper is to explore concepts of information that have been suggested, and to propose a definition that encapsulates them. It is not intended to present new ideas, but rather to consolidate existing ideas in a way that makes them easy to put across to students. The complexity of ideas concerning information is such that a paper this length can scarcely do them justice, but it is intended that this paper should provide a starting point to which students can relate modules as diverse as document retrieval and systems modelling.

What is information?

Attempts to answer the question 'What is information?' have, not surprisingly, occupied the thoughts of information scientists for a long time: almost certainly since before the term 'information science' was coined in 1955 [3]. The lay person, asked to define information, is most likely to regard it as:

> An item of information or intelligence; a fact or circumstance of which one is told. (OED)

This is just one of the many dictionary definitions of the word. Indeed, information scientists appear to have been reluctant to propose definitions of information, preferring rather to discuss concepts the difference being, according to Belkin [4, p. 58], a definition 'says what the phenomenon defined is, whereas a concept is a way of looking at or interpreting the phenomenon'.

In their recent paper, McCreadie and Rice [5] review concepts of information proposed over the last fifty years. A summary of the concepts they consider is given below.

- **Information as a representation of knowledge**
  Information is stored knowledge Traditionally the storage medium has been books, but increasingly electronic media are becoming important.
- **Information as data in the environment**
  Information can be obtained from a range of environmental stimuli and phenomena; not all of which are intended to 'convey' a message, but which can be informative when appropriately interpreted.
- **Information as part of the communication process**
  Meanings are in people rather than in words or data. Timing and social factors play a significant role in the processing and interpretation of information.
- **Information as a resource or commodity**
  Information is transmitted in a message from sender to receiver. The receiver interprets the message as intended by the sender. There may be added value as the information is disseminated or exchanged.

Information in context

The model presented below rests on the assumption that information cannot be evaluated without an awareness of the context in which it is being interpreted. This assumption leads to a model comprising three components.

1. Readership context

The context in which a message is received and interpreted. The reader is any system which derives (or attempts to derive) information from a message. A system may be a mechanism, an organism, a community, or an organisation.

2. Authorial context

The context in which the message originates. The author is any system that transmits (intentionally or otherwise) a message from which a reader can derive information.

3. Message

The means by which information is transmitted. It may be written, spoken, facial expression, pheromonal, etc.

These components are described more fully below.

Information as data in the environment: reading the signs

Many information scientists accept that information is a property of all living organisms [6, 7]. It is not unreasonable, therefore, to illustrate the prime importance of context with an example from biology. It has long been recognised by zoologists that there is an association between brightly coloured markings on an animal and unpalatability or toxicity [8]. Predators learn to associate such markings with unpleasant experiences and so are less inclined to attack similarly marked creatures in future. So a potential predator, seeing the markings of a brightly coloured male butterfly, will derive information about the insect's suitability as a food source. A female butterfly of the same species will derive no such information. She might, however, find the markings useful in assessing the male's quality as a mate. Clearly, therefore, both the predator and the female butterfly derive information from the markings, but the message of the markings depends on the context in which it is read.

The importance of context in the interpretation of information has long been recognised. The above example, however, indicates that, unless it is assumed that butterflies and birds have knowledge, knowledge is not necessary for a signal to be informative.

Information as part of the communication process: the authorial and readership contexts

Messages exchanged between humans frequently differ from those discussed in the two examples above because there is often an intention that they should carry information. There is no reason to assume that a male butterfly means to inform either a predator or a potential mate when he flaps his wings. By contrast, a message designed to communicate has two informing contexts: that of the author and that of the reader. These correspond to the two points described by Shannon and Weaver [9, p. 31] when they stated that 'The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.' However, Shannon and Weaver were merely talking about the difficulties involved in transmitting a signal from a sender to a receiver. If that signal is to be a message, it is necessary for the sender to be an author, or the recipient to be a reader, or both. Characteristics of these two contexts, and the message itself, are described below.

The readership context

As Meadow and Yuan noted, 'Most views of the difference between data and information ... depend on the recipient.' [10, p. 701] The information derived from a message by a reader depends on a wide range of factors, all of which affect the reader's understanding of that message. Some of these are listed below:

- Geographical – nation, culture, language, physical community.
- Social – interests, pastimes.
- Educational – level of education, subjects studied.
- Professional – area of professionalism, career history.

The different contexts overlap. A mathematical treatise will be understood in the same way by both Russian and American mathematicians. A Birmingham newspaper will be more informative to Jamaican and Punjabi immigrants living in Solihull than it would to a tenth generation cockney in Lambeth. To understand what makes sense to a reader, therefore, it is necessary to understand the structure of the society of which he or she is a part [11].

The authorial context

As well as sharing the characteristics of the readership context, the authorial context has an additional property: that of intention. Two possible states of intention are assumed:

Message intended to convey information
The author produces the text with the intention of informing the reader. This is the usual authorial context, in which a text 'is a collection of signs purposefully structured by a sender with the intention of changing the image-structure of a recipient' [12, p. 20]. The closer an author's context is to that of a reader, the greater is the chance that the author's work will be informative. In exceptional circumstances, an author may choose to convey more than one message in a given text (see Appendix).

Message not intended to convey information
The author ascribes no meaning to the message of the text: any meaning is derived within the readership context. Examples include the predictions of fortune-tellers and output from artificial intelligence programs such as Eliza [13].

Information as a resource or commodity: getting the message

Messages are classified according to how focused they are. Traditionally, information scientists have dealt largely with focused messages. Highly focused messages are ones in which the context for interpretation is very specific, making ambiguities difficult or impossible. The most obvious example would be a mathematical document, but other examples include command line computer interfaces and technical publications.

Looser (but still focused) messages would include descriptive works and histories, which will be interpreted according to the reader's culture and experience. A less positive example would be poorly written documents [4], which may be confusing, ambiguous, or misleading: a typical contemporary example of this is email. Totally loose messages would have no obvious interpretation in any context. An example is surrealist literature. In many ways this model is similar to the communication model proposed by Jakobson [14]. This too comprises three components (addresser, addressee, and message). Because it deals with the intentional transfer of information, however, it places greater emphasis on the means by which messages are transmitted, and excludes a great deal of information sources.

The importance of recognising context

Wilson [3] stresses the importance of context in dictating information needs, but makes no mention of the impact of context on the interpretation and effectiveness of information materials. Hjørland considers 'subject analysis of documents as one of the most fundamental activities of library and information professionals' [2, p. 610].

One implication of the model presented here is that such an analysis requires an appreciation of the context. As Hjørland notes:

> The subject of a book (or any other document, or message) is closely related to what kind of answers people can find from reading the book... Any document thus has an infinite number of subjects [2, p. 610].

So to predict the effect of information on a particular user, it is necessary first to envisage the potential user. In classifying information, therefore, the information scientist is implicitly classifying the user.

Information and knowledge

Earlier, the importance of context in causing a stimulus to become informative was discussed, but arguably the most fundamental context was omitted: that of the knowledge of the recipient of the information. The examples of contexts listed above will all shape that knowledge, hence their significance, but in addition the knowledge will be affected by an incalculable array of experiences and aptitudes.

The association between knowledge and information seeking is well established: '...the idea of using cognitive models as the basis for information retrieval system design has aroused considerable interest...' [15, p. 63].

It has been argued that what motivates someone to seek information is a recognition by the seeker of 'an anomaly in his/her state of knowledge' [4, p. 81]. Moser [16, p. 350], questions how 'normal' and 'anomalous' states of knowledge are to be identified or measured, and makes the point that 'information, to be generated, need not actively be instigated on the 'recipient's' side'.

This observation is highly relevant to the context-reliant model of information reception described in this paper. Checkland argues that 'consciousness makes man, via his W(orld View)s, a meaning-endowing animal' [17, p. 219]. If this is accepted, then it is the meaning endowed within the World View that will determine whether information is sought, what information is sought, and how it is interpreted. Anomalies may be a motivating factor. People who, unlike Lewis Carroll's 'Humpty Dumpty', are not happy to 'believe six impossible things before breakfast' may seek information in an effort to resolve some of the discrepancies in their World View.

Alternatively, however, the meaning conferred by a World View may provide someone with a paradigm of 'normality', which can be used in the generation of hypotheses. Here, information may be sought to test the hypotheses in order to establish or to extend 'normality'.

Personal paradigms as context

The word 'paradigm' above is used in its dictionary sense: 'pattern, example, to exhibit beside, show side by side' (OED) rather than in the ways in which Kuhn [18] used the term in his philosophy of science. While Kuhn used 'paradigm' to describe systems by which meaning could be shared in a research community, the word is used above to describe the ways in which an individual organises information within his or her World View.

The difference is significant when considering the evaluation of information. Information scientists can only assess information insofar as their World Views match those of the people for whom they are evaluating it. It is because the match is inexact that browsing and serendipity are important factors in information seeking. Data and texts that appear irrelevant to an evaluator may provide the missing piece of a puzzle to a researcher; but the data and texts will only be informative if the puzzle is known.

Probably the best known example in science of a serendipitous discovery arose because of the problem of King Hieron's crown. The king, wishing to know whether the crown was pure gold as claimed by the goldsmith, or whether a gold/silver alloy had been used, asked Archimedes to investigate. Archimedes is reputed to have arrived at the solution when, as he climbed into his bathtub, he observed water overflowing from it. Within the context of his knowledge, his thoughts, and his ideas, the stimulus of overflowing water was informative. Archimedes deduced that the quantity of water displaced was equivalent to the volume of his body, and so had a means of determining the density of the crown [19].

The history of science is full of such tales: from the apple that gave rise to Newton's thoughts on gravity, to the dream of snakes from which Kekule derived the structure of benzene. Such examples, however, are of little practical relevance to the information scientist, since information of this kind is impossible to organise. A more constructive and more recent example involves the work of Heisenberg in quantum mechanics. According to C.P. Snow, in the early 1920s, Heisenberg was seeking to find mathematical tools which would enable him to relate the set of rules associated with any given atom to that atom's set of properties.

> The trouble was, he didn't know enough of the curiosities of nineteenth- century mathematics, when all kinds of mathematical arts had been developed. Not for use, but for the sheer beauty of the game.
>
> Fortunately... Max Born ... [knew of the] old subject of matrix algebra, half forgotten but completely available [20, p. 67].

This half-forgotten branch of mathematics proved to be 'precisely what they needed'. If Snow's analysis is correct, however, and matrix algebra had indeed been developed for 'the beauty of the game' rather than for use, it would presumably have been held to have little informative value. Despite this, the text was 'completely available' and clearly retrievable; and within the context of Heisenberg's research it became invaluable information.

Information in the community

Although the knowledge of the information user may be the ultimate informing context, from the point of view of an information manager wishing to provide relevant information it is probably an impractical starting point. It is for this reason that information specialists tend to deal with information as a representation of knowledge, or 'information-as-thing' [21].

As was stated above, however, community is also an important context. For the purposes of the information manager, this is usually the context by which information is defined. What is stored in collections of informative things, whether those collections be archives, libraries, or digitised records, is stored with a view to the needs of the target community.

The idea that information is embedded in socio-cultural contexts is not new of course [22]. What has changed for the information professional, however, is the extent to which the nature of the community being served must be considered. The traditional librarian was usually a professional, catering for other professionals educated to a similar level, and therefore well able to anticipate their requirements. In the case of public libraries, the user community defined itself by its desire to use the library. This is still true, but given the ever-expanding range of alternative sources of information and entertainment, this community is declining.

Memes and the transfer of information

As was stated at the start of this paper, information is widely regarded as being a property of living organisms. Dennett in particular, stresses the connection between information and awareness in many forms of life [23], and discusses the impact of information on consciousness. In so doing, he draws heavily on some of the ideas proposed by Dawkins [24] in *The Selfish Gene*. Here, Dawkins argues that a lot of ideas are reproduced in human society in a manner analogous to genetic replication. He refers to such ideas as memes, and the resulting study (mimetics) has begun to gain acceptability. Mimetics draws heavily on comparisons with biological evolution, and the succession of overlapping contexts described above bears a similarity to Hutchinson's classic definition of an ecological niche [25] as an n-dimensional hypervolume: a mathematically defined space in which each of the factors affecting the viability of an organism occupying that niche is seen as a separate dimension.

It has been commonly observed that 'Almost always the men who achieve... fundamental inventions of a new paradigm are either very young or very new to the field whose paradigm they change.' [18, p. 90]

If the ideas presented above on community as a informing context are 'linked' to those on World View, the possibility is raised that what is learned in one community will, in the context of a different community, be informative in ways that were not previously recognised. To extend the evolutionary analogy use in mimetics, this perhaps represents a cross-fertilisation of ideas.

Conclusion

As has been argued, the materials with which the information scientist routinely works represent just a small proportion of potential information. This paper therefore proposes that, to capture the 'breadth' of possible information sources, information should be defined as:

> a stimulus originating in one system that affects the interpretation by another system of either the second system's relationship to the first or of the relationship the two systems share with a given environment
> (where a system is as defined above, in Readership Context).

Brown [1, p. 185] suggests that, in attempting to define information, information scientists have tended to restrict the term to only a part of the whole and that 'no satisfactory concept of information for information science will ever be formulated in the sense of supplanting all others'.

The model described above represents an attempt, not to supplant concepts of information, but to unite them. The idea that information is only information in certain 'informing contexts' incorporates nearly all widely held concepts of information, and would provide a useful point at which to introduce students to the range of directions in which information science could take them.

Appendix

Deliberate ambiguities are common in puzzles, codes, and as literary devices. Simple codes may carry both overt and cryptic messages. For example, the message:

> Coming in tomorrow evening. Meet at
> Dinnington depot, eleven nineteen.

would convey to most readers the information that a liaison is being requested. Those able to apply an appropriate geographical context will derive additional information concerning the location. This instruction may be relevant to the intended recipients of the message, but they would also receive further information by being aware that they should read the initial letters of the words. Punning headlines are commonly used to summarise two aspects of a newspaper story. A fictional example would be that of John Smith, an aspiring rock star arrested for assault. A report on his chart success and his appearance in a police identity parade may be headlined: 'SMITH IN HIT PARADE'. Such exercises need not be limited to one language. A French-speaking reader of the book *Mots d'Heures: Gousses, Rames* (Verse 11) [26] will struggle to understand esoteric and surreal verse such as:

> *Chacun Gille*
> *Houer ne taupe de hile*
> *Tôt-fait, j'appelle au boiteur*
> *Chaque fêle dans un broc, est-ce crosne?*
> *Un Gille qu'aime tant berline à fêtard.*
> (Verse 11)
> *(Every bumpkin*
> *While hoeing uncovers a mole and*

part of a seed.
Quickly finished, I call to the limping man that
Every pitcher has a crack in. It is it a Chinese cabbage?
A bumpkin loves a life of pleasure and a carriage.)

An English listener, however, hearing the verse read aloud in French, will find it 'hauntingly familiar'.

References

1. Brown, A.D. *Towards a theoretical information science: information science and the concept of a paradigm.* (Department of Information Studies, Occasional Publication No. 5) Sheffield: University of Sheffield, 1987.

2. Hjørland, B. Theory and metatheory of information science, *Journal of Documentation,* 54 (5), 1998, 606-621.

3. Wilson, T.D. Models in information behaviour research. *Journal of Documentation,* 55(3), 1999, 249-270.

4. Belkin, N.J. Information concepts for information science. *Journal of Documentation,* 34(1), 1978, 55-85.

5. McCreadie, M. and Rice, R.E. Trends in analyzing access to information. Part I: cross-disciplinary conceptualizations of access. *Information Processing and Management,* 35(1), 1999, 45-76.

6. Debóns, A. and Horne, E. NATO advanced study institutes of information science and foundations of information science. *Journal of the American Society for Information Science,* 48 (9), 1997, 794-803.

7. Vickery, B. Metatheory and information science. *Journal of Documentation,* 53(4), 1997, 457–476.

8. Ruxton, G. Sheep in wolves' clothing. *Nature,* 394, 1998, 833–834.

9. Shannon, C.E. and Weaver, W. *The mathematical theory of communication.* Chicago: University of Illinois Press, 1949.

10. Meadow, C.T. and Yuan, W. Measuring the impact of information: defining the concepts. *Information Processing and Management,* 33(6), 1997, 69-714.

11. Dervin, B. On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing and Management,* 35(6), 1999, 727–50.

12. Belkin, N.J. and Robertson, S.E. Information science and the phenomena of information. *Journal of the American Society of Information Scientists,* 27(4), 1976, 197–204.

13. Allen, J. *Natural language understanding.* 2nd edition Redwood City, CA: Benjamin/Cummings, 1995.

14. Jakobson, R. Closing statement: linguistics and poetics. In: Sebeok, T., ed. *Style and Language.* Cambridge, Mass.: MIT, 1960, 350–377.

15. Ellis, D. *Progress and problems in information retrieval.* 2nd edition London: L.A. Publishing, 1998.

16. Moser, A. Information concepts. *Journal of Documentation,* 34(3), 1978, 350–1.

17. Checkland, P.B. *Systems thinking, systems practice.* 2nd edition Chichester: Wiley, 1984.

18. Kuhn, T.S. *The structure of scientific revolutions.* 3rd edition Chicago: University of Chicago Press, 1996.

19. Ronan, C.A. *The Cambridge illustrated history of the world's science.* Cambridge: CUP, 1983.

20. Snow, C.P. *The physicists: a generation that changed the world.* London: Macmillan, 1981.

21. Buckland, M.K. Knowledge as thing. *Journal of the American Society for Information Science,* 42(5), 1991, 351–360.

22. McGinn, C. *Wittgenstein on meaning: an interpretation and evaluation.* Oxford: Blackwell, 1984.

23. Dennett, D.C. *Consciousness explained.* London: Penguin, 1993.

24. Dawkins, R. *The selfish gene.* Oxford: Oxford University Press, 1989.

25. Hutchinson, R.E. Concluding remarks. Cold Spring Harbor. *Symp. Quant. Biol.* 22. 1957, 415-427.

26. Van Rooten, C.H.K., *Mots d'Heures: Gousses, Rames.* UK: Angus & Robertson Ltd, 1968.

What should we understand by information technology (and some hints at other issues)?

Julian Warner
School of Management and Economics, The Queen's University of Belfast, Belfast BT7 1NN
j.warner@qub.ac.uk

Information science has been convincingly characterised as a response to developments in information and communications technologies and as part of the gestalt of the computer. Despite this, it has had a limited understanding of information technology and has repressed or disguised its origins. Its understanding of itself and its potential for contribution to other discourses has thereby been restricted.

The paper develops an understanding of information technology. The idea that the computer as a machine is concerned with the transformation of information, not material or energy, is extended to other information technologies. Technology is regarded as a radical human construction, in a position derived from Marx and mediated by economics. On these bases, an understanding of information technology as a form of knowledge concerned with the transformation of signals from one form or medium into another is proposed. Invention, innovation, and diffusion are distinguished as stages in the development of technologies.

For modern information technologies, the history of copyright can provide indicators for innovation and diffusion. The mid- to late 19th century, in the United States and between the United States and Europe, is identified as the critical period for diffusion. An explanation for this is proposed in terms of the dynamism of the period, its hospitality to innovation, and in the United States continental expansion and developing links with Europe.

Introduction

From the thesis that being or social existence determines consciousness, the set of activities labelled as information science [1] have been convincingly regarded as a partly unreflexive response to developments in computer and telecommunications technology from the late 1940s [2; 3, pp. 113–115]. Specifically, information science has been understood as part of pattern where 'New tools beget new sciences' [3, p. 115]. Yet information science has had a limited understanding of information technology and could be regarded as having repressed or disguised its origins.

The understanding of information technology has been limited in a number of ways, often by implication rather than by explicit statement. Information technology may be regarded as objectively given or an autonomous development, particularly as implied by the language of discussion [4, 5]. There are also traces of technological determinism, most subtly and pervasively in the limited recognition of information technology as a human construction [6, 7, 8, 9]. Historical consciousness has also been limited and the crucial question of the critical period for the development of modern information technologies has only been briefly addressed [10]. The positions implied, and here identified, may not reach the level of formal articulation and can appear weak when articulated. A specific feature of repression would be to claim to focus on information processes but then to allow topics for investigation to be largely determined by the potentialities of current information technologies. These assertions would admit of qualification, particularly with regard to the growth of historical consciousness, and have been partly eroded by recent contributions, but still characterise the dominant tenor of the field.

There are some analogies with extra-disciplinary perspectives, which may have also influenced intra-disciplinary thought. A strong, although largely implicit, analogy with the reception of technology as objectively given would be with the, often dominant, Platonist view of mathematical propositions as discovered rather than humanly created. This may have unconsciously, but still powerfully, have influenced the thinking of information scientists from mathematical and scientific backgrounds [1]. More broadly, there has been a divide of technology from social and human thought: crudely put, technology has been left to the technicians, who may not be concerned with issues of human and social significance [11]. One strong humanist attitude has been to regard technology as something artificial, a perspective plausibly inherited from responses to industrial technology. The linguistic resources of public discourse remain inadequate to understanding current technology [12; 13, pp. 2–11]. An unexplored socio-technical dynamic persists [14, p. 1087].

The limited understanding of information technology in information science is unfortunate, in two major respects. It restricts its understanding of itself, of its historical development, and the historical circumstances that produced it. It also limits its contribution to other disciplines and to wider public discourse, although there are elements in information science, and in contiguous and contributing areas of study, which could contribute to a theoretically sophisticated and historically specific understanding of information technology.

The aim here is to begin the task of articulating a counter-position to the limited understanding of information technology in information science. Drawing on a contributing and partly contiguous discipline, on theoretical computer science, the conception of the computer as a machine will be expanded to encompass other information technologies. The concept of technology, elided in the understanding of the computer as a machine, will be reinvigorated by drawing on an understanding of technology as a human construction, derived from Marx and largely mediated by economics. The further developments envisaged on the basis of the understanding of information technology established will be briefly foreshadowed: the series of systematic distinctions, from invention, through innovation, to diffusion, which can be used to explore the socio-technical dynamic, will be introduced; evidence for the late 19th century United States as the critical period for the development of modern information technologies will be reviewed; and the question of why these technologies should have been produced at that time and in that place will be addressed.

Computer as a machine

A sophisticated discussion of the computer as machine is offered by Minsky [15]. In this context, it is valuable both for distinctions made and for distinctions implied, with a continuum between these.

The most crucial distinction is explicitly made in a differentiation of the computer from productive technology. The computer does not:

> do 'work' as did machines of the Industrial Revolution. Instead of dealing with material or energy ... they handle 'control' and 'information' [15, p. vii].

Minsky himself notes that the classical idea of a simple machine, such as a lever, wheel, or inclined plane, only helps understand the transmission of force and not the control mechanisms within such machines, such as the ratchet [15. p. 7]. The distinction from productive technology, and the concern of the computer with transforming information, has been accepted in computer science, at the level of a grounding assumption. The proposal here, which should be equally readily acceptable, is explicitly to extend this distinction and conception to other information technologies. At the initial level of analysis, no distinction is made between technologies used in human communication and control mechanisms within machines, and, for some purposes, this would be a valuable further distinction.

The image of the machine is also considered. The use of the term machine in ordinary discourse does not evoke a sympathetic object:

> It brings to mind a big, heavy, complicated object which is noisy, greasy, and metallic; performs jerky, repetitive, and monotonous motions; and has sharp edges that may hurt one it he does not maintain sufficient distance [15, p.1].

Notions of mechanism and determinism may be simultaneously evoked [15, p. 1]. The implication of an inheritance form and continuity with attitudes to industrial technologies can be briefly contextualized. For instance, Charles Dickens in *Hard Times* was said to view Coketown and its industrial machinery 'as a slightly disgusted Southern visitor would see it' [16, p. 48]. Similarly, Norbert Wiener comments on the image of the shed factory as 'one of the crueller circles of the hell of the Industrial Revolution' [17, p. 93]. The unsympathetic image of the machine may partly account for the reluctance to acknowledge technology as a human construction. In addition, the association of both productive and information technology with mechanism and determinism might further explain the humanistic resistance to technology, currently being eroded [18].

A further description of the computer offered by Minsky, as a strange new species [15, p. viii], is more valuable for what it implies or reveals than what it explicitly states. It anticipates the widely promulgated view of the computer as a 'radical novelty' [19, 20, 21]. In this context, it can be regarded as an unconscious reproduction of the innovatory dynamism of capitalism, with its revolutionary rather than conservative technical base [22, p. 617]. It also obscures the human labour involved in the production of the computer, and, as has been more widely recognised, its historical antecedents in other information technologies [20]. At the level of specific historical events, rather than theoretical understanding, Minsky recognises the historical development of the theory of computation and of the computer and questions the received view of a divorce between seminal papers in mathematical logic in the 1930s and the construction of working computers [15, p. 104]. The received view does reinforce the separation between the understanding of the possibilities and limits of computation and of the computer as a form of technology.

The aim of the investigation and exposition of the computational process is, for Minsky, to develop a grasp of fundamental matters and bring this understanding back to the practical world [15, pp. 2–3]. With some exceptions, the challenge to return to the practical world has not been accepted and automata studies has been pursued as an autonomous activity, with a curious combination of restrained theory and mystical practice observed [20], and, possibly, resulting [21]. The proposal here is to extend the differentiation of the computer from industrial machines, as being concerned with the transformation of information inputs into information outputs, to other information technologies and to reinvigorate the understanding of technology.

Understandings of technology

The classic source for an understanding of technology as a human construction is Marx, although his work has been subject to neglect and distortion [23; 24, p. 10]. Since the early 1990s, Marxism has become intensely unfashionable [25, p.18]. For earlier periods, it has been convincingly suggested that the strong polemical orientation associated with Marxist studies has limited the diffusion of its treatment of technology [26, p. 126]. A misleading view of Marx as technological determinist did, however, gain sufficient credence to be worthy of repudiation [25, pp. 23–46]. Marx's views on technology have been largely mediated by economics, particularly by Rosenberg [26, 27, 28], and although information technologies, such as the telegraph, have been studied [28, pp. 203–231], the issue of their differentiation from productive technologies has not been explicitly addressed. With the exceptions of Wiener, who both by explicit allusions and thematic echoes reveals an acquaintance with Marxist thought [17, p. 4; 29], and, more obviously and polemically, J.D. Bernal, there has only been a limited penetration of Marxism into information science.

Yet, as in economics, with its primary focus on productive technologies, a Marxian perspective on information technology could be of high marginal utility.

The social construction of the technology movement is better, although not fully, known with information science, particularly through recent imports [30]. A comparative reading would suggested that it has developed partly, although not entirely, separately from the Marxian approach mediated by economics, and is not necessarily fully aware of potentially relevant antecedents. Its focus has been on computer technologies, and, within that, on software rather than on the computer as a machine. To some extent, it has also emphasised the social shaping rather than the radical human construction of technologies [31, p.270]. The Marxian approach promises to be richer and fuller, although, as yet, less fully developed, inclusive of productive and other information technologies, of the computer as a machine as well as software, and emphasising the human construction as well as the social shaping of technologies. The intersection, or lack of intersection, between the Marxian approach mediated by economics and the social construction of the technology movement could form a focus for further study, possibly supported by bibliometric methods, with a view to their mutual synthesis.

A view of technology as a human construction is given by a classic passage in the *Grundrisse*:

> Nature builds no machines, no locomotives, railways, electric telegraphs, self-acting mules etc. These are products of human industry; natural material transformed into organs of the human will over nature, or of human participation in nature. They are *organs of the human brain, created by the human hand*; the power of knowledge objectified. The development of fixed capital indicates to what degree general social knowledge has become a *direct force of production*, and to what degree, hence, the conditions of the process of social life itself have come under control of the general intellect and been transformed in accordance with it. To what degree the powers of social production have been produced, not only in the form of knowledge, but also as immediate organs of social practice, of the real life process [32, p. 706].

In accord with the tenor of the industrial age [10], the primary emphasis of this passage is on technology as a means for transforming the material environment, and, although transport and information technologies (both for human communication and machine control) are cited, they are not explicitly differentiated.

Some further aspects of the passage deserve comment. The crucial distinction, which guides Marx's treatment of technology, is made between natural developments and human products. From this perspective, even domestic animals would be regarded as a form of technology as they have been deliberately modified by human intellectual and physical labour, *malgré eux* [22, p. 285]. Knowledge as social practice is distinguished from knowledge in itself. In the reference to the 'general intellect', there is an anticipation of categories later explicitly introduced in the third volume of *Capital*, of universal and communal human labour:

> They both play their part in the production process, and merge into one another, but they are each different as well. Universal labour is all scientific work, all discovery and invention. It is brought about partly by the co-operation of men now living, but partly also by building on earlier work. Communal labour, however, simply involves the direct co-operation of individuals [33, p. 199].

The distinction can be discovered in domains familiar to information science: patent rights could be regarded as a restriction on the diffusion of the products of communal human labour and the transition into the public domain as a movement from communal to universal human labour. As universal labour, science and technology are viewed as cumulative processes, which may then have some autonomy from the immediate economic base [3, p. 115]. A distinction between objectification and reification would also be relevant here: objectification would refer to the process of the transformation of knowledge into 'organs of social practice ... the real life process'; reification to allowing the thing like nature of technology to disguise the universal and communal labour embodied in it. Where information science has received technology as objectively given or as an autonomous development, it has fallen into the false consciousness of reification. The crucial points to be taken from this passage are, then, that technology is a human construction, not a natural development, made historically and cumulatively, by human labour on nature, and that the materiality of technology is not denied but seen as the 'the power of knowledge objectified'.

An example may help understand these themes and appreciate the insights the perspective can yield. Consider, for instance, the contrast between the replication of the work of British novelists in the late 19th century United States, without copyright being accorded to the authors, and the distribution and reproduction of the Starr report. For the 19th century works, copies were physically transported from Britain to America and then, in the case of the 'cheap books' of the 1880s, reproduced by being reset by non-unionised labour, without proof-reading (although the trans-Atlantic cable (1866) preceded the recognition of international copyright (from 1891), it does not seem to have been used for the transmission of novels, presumably because of the time and cost of signal production and transmission) [34]. The Starr report, in contrast, was first distributed electronically, with some reports of delays in signal transmission, and reproduced at the point of reception, for reading. Replication of the report as a published printed book, available for purchase, rapidly followed. There are familiar contrasts in the speed of dissemination and the effects this has on the significance of space and place (a concept of communication distance can also be found in Marx [24, pp. 136, 151, 156]. Beniger, for instance, regards speed as the most significant parameter of the control revolution [35, p. 175]. Precise numerical comparisons of speeds of transmission and replication may not yield great insights, but there may be a greater contrast in the pace of transmission, with physical transport replaced by signal transmission, than in the time required for reproduction, with automatic processes displacing human labour in typesetting. Far less noticed, but emerging forcibly from the perspective here, is the reduction in the quantity and skill of human labour required at the point of replication. A degree of historically accumulated universal labour and of communal labour was embodied in the late 19th century printing technology; greater physical and, particularly, intellectual labour, was embodied in late 20th century communications and information technology; accordingly, less skill and understanding, and a lesser amount of labour, was required at the point of replication. The historical reduction of living labour required also exemplifies the fundamental dynamic of capitalism, the substitution of dead for living labour, although with regard to signal transmission and data transformation rather than the transformation of natural resources.

The distinction between objectification and reification can also be understood in relation to these examples. The 'cheap books' (now rare) of the 1880s reveal their process of production by misprints and poor typography; analogously, delays in signal transmission indicated the humanly constructed nature of the Internet. For Marx:

> it is by their imperfections that the means of production in any process bring to our attention their character of being the products of past labor. ... In a successful product, the role played by past labor in mediating its useful properties has been extinguished [22, p. 283].

A strong analogy can be made with the reconstruction in analytic bibliography of the diachronic filiations and the synchronic circumstances of production of a work from the transmission and commission of errors. Similar techniques have been used to detect citation of references from secondary sources. In these instances, the imperfections in the final products reveal their processes of production, bring the human labour involved in their making to attention. Reification would be more likely to occur when the finished appearance of the product disguises the labour embodied in it.

The view of technology as a human construction developed in the *Grundrisse* underlies the proposal for a history of technology given in *Capital*:

> A critical history of technology would show how little any of the inventions of the eighteenth century are the work of a single individual. As yet such a book does not exist. Darwin has directed attention to the history of natural technology, i.e. the formation of the organs of plants and animals, which serve as the instruments of production for sustaining their life. Does not the history of the productive organs of man in society, of organs that are the material basis of every particular organization of society, deserve equal attention? And would not such a history be easier to compile, since, as Vico says, human history differs from natural history in that we have made the former, but not the latter? Technology reveals the active relation of man to nature, the direct process of the production of his life, and thereby it lays bare the process of the production of the social relations of his life, and of the mental conceptions that flow from those relations. Even a history of religion that is written in abstraction from this material basis is uncritical [22, pp. 493–494].

The proposal has been regarded as a prolegomenon to a history of technology, which had yet to be written [27, p. 34]. Marx's concern is with forms of technology concerned with the human transformation of the material environment, but the extension of this perspective to information technologies, as a component of an information view of history, has been proposed [10]. Such a history remains even less written, but this paper is itself intended as a contribution.

Some aspects of the passage again deserve comment. The view that mental conceptions flow from social relations can be regarded as an alternative formulation of the theme that being determines consciousness. The mediating role given to technology as the 'basis of every particular organisation of society' need not be regarded as technologically determinist, if the technology itself is regarded as a human construction, produced and adopted under specific historical circumstances. A significant critique of an emphasis on technique determining work relations further clarifies this position: for Marx, the 'purely technical aspect is seen only as moment of the overall socio-economic process'[36, p. 55n]; and the possibility of '*reciprocal interaction* by no means suppresses the real historical and methodological primacy of the economy over technique' [36, p. 56]. The reference to technologies as not the work of single individuals recalls the stress on the social intellect in the *Grundrisse* and anticipates the later categories of universal and communal human labour.

The reference to Darwin and natural technology does indicate the possibility of an evolutionary view of technology as a cumulative construction [27, p.6], but also exposes a crucial contrast. The 'organs of plants and animals' are natural developments, whereas technology is a human construction, with domesticated animals as an example of human work upon nature. Humanly created technology is also revealed as equally important as 'natural technology'. The view of work, of productive labour on nature, as condition of fallen human existence is reminiscent of *Genesis* and does place Marx in the Judeo-Christian tradition. The formula the men make their own history, here attributed to Vico, is central to Marx's thought. A subsequent notion in studies of technology, of path-dependency, could be regarded as exemplifying the qualification on this position, 'but not of their own free will; not under circumstances they themselves have chosen but under the given and inherited circumstances with which they are directly confronted' [37]. Vico might be more familiar from his view that the axioms as well as the subsequent propositions of mathematics are to be regarded as human constructions, not received as objects of contemplation [21; 38, pp. 50–51], and does reinforce the possibility of discovering extensive analogies between critiques of technology and of mathematics or logic as objectively given.

The crucial points to be taken from this passage are the possibility of the historical study of technology as a cumulative human construction and the significance of those technologies to the production of social relations. For Marx, writing, in the 1860s, it was only in large-scale industry that man had made objectified labour 'perform gratuitous service on a large scale' [22, p. 51]. From an information science perspective, information technologies can be regarded as comparably significant to industrial technologies to 'every particular organization of society' [10]. Information technologies, developing extensively since the date of *Capital*, can then similarly be historically studied as cumulative human constructions.

On the basis offered by Marx, partly focusing on the crucial passages considered from the *Grundrisse* and from *Capital*, a modern understanding of technology has been constructed, particularly by Rosenberg [26, 27, 28, 39]. Citation data simultaneously testifies to the influence of his work and to its relative lack of importation into information science. The diversity and complexity of technology is taken as axiomatic [26, p. 2]. Yet a degree of unity is also perceived:

> the most useful common denominator underlying its [technical progress] multitude of forms is that it constitutes certain kinds of knowledge that make it possible to produce (1) a greater volume of output or (2) a qualitatively superior output from a given amount of resources [27, p. 3].

In an alternative formulation, technology is regarded as form of knowledge concerned with the transformation of natural resources into useful goods and services. The obdurate materiality of the machine is not denied but regarded as essential to its social function [25, p. 9]. In terms of the distinctions established, objectification, the transformation of knowledge into part of the real life process, is admitted, but reification, denying the human labour embodied in technology, avoided. The difficulty of modelling technical change is conceded and it is recognised that only a rudimentary understanding of society's capacity for generating progress has been developed [27, p. 29]. In particular, the possibility of a qualitatively superior output, and not just its quantitative development, is significant but has not been sufficiently recognised by economics [27, pp. 3–4].

Information technology can be assimilated to this model. Drawing on the distinction derived from the discussion of the computer as a machine, it can be regarded as a form of knowledge concerned with the transformation of information inputs into information outputs, or, with potentially greater terminological clarity, with the transformation of signals from one form or medium into another. The understanding proposed deserves isolation and repetition:

> *A form of knowledge concerned with the transformation of signals from one form or medium into another.*

Again, at the first level of analysis, no distinction is made between the use of technologies in human communication and in machine control. Where the concern is with the complex technical antecedents to modern information technologies [40, p. 699], it may be helpful to treat the field in this unified, and undiscriminating, way; where the focus is specifically on human communication or on control, it would be helpful to introduce the further distinction (a distinction by function). The concept of technology has, then, been reinvigorated and discourse about information technologies productively linked to, but still differentiated from, discourse about other technologies.

A specific connotation is intended by *signals*, related to classic distinctions made in information theory, information science, and in semiotics. In *The mathematical theory of communication*, the meaning of symbols was not relevant to considerations of the economic and efficient transmission of messages. A deliberately restricted sense of *information*, excluding meaning, was used. At the same time, optimism was expressed that the theory developed would be relevant to the understanding of human communication at the semantic level [41], and, this, together with the ordinary discourse connotations of information, has been a source for subsequent confusion. Bar-Hillel, in a well-known critique, reiterated and emphasised the distinction present in *The mathematical theory of communication*, characterising it as theory of signals not information [42]. More recently, in information science, information as thing has been differentiated from information as process and information as knowledge [43]. Signals would carry the connotation of information as thing. An analogous distinction has also been made between dissemination, or message transmission, and diffusion, the growth of understanding, and again, signals would correspond to the level of dissemination [44]. In the rigorous, and potentially clarifying, terminology of Saussurean semiotics, the concern is with the signifier or expression (the term *signal* has itself been used for this aspect of the sign, in an English translation of Saussure [45]), not with the signified or content, nor with the sign, the union of signifier and signified [21]. Advantages in terminological clarity could be derived from the use of *signals* in preference to *information* and this move will be followed here.

Some attention should also be given to the understanding of knowledge, drawing partly on sources more familiar to information science and opening up the possibility of a reciprocal contribution to studies of technology. A distinction should be made, methodologically at least, between forms of knowledge open to symbolic, including verbal, articulation (and a greater degree of exactness may be demanded by written rather than oral forms of verbal expression [46]) and forms of knowledge resistant to reduction to symbolic form.

Even with forms of knowledge open to symbolic articulation, it would be wrong to identify the knowledge with its representation in a fixed text. Knowledge exists in a state of perpetual movement. To deny this is to exemplify it. There may be forms of knowledge transmission, for instance, didactic teaching [47], and forms of knowledge representation, such as the finished nature of a published book, which give it the appearance of relative fixity. In some discussions of the philosophy of mathematics, the interpretative assent of the relevant community is required to transform a written text into accepted knowledge. Analogously, in semiotics human interpretation is required to create signs from signifiers [21]. Wiener distinguished between the texts published by scientists and their continual creation and re-creation of knowledge, memorably observing, 'There is no Maginot Line of the brain.' [29, p.122] An analogy could be made with the living labour needed to quicken the dead labour embodied in machinery [22, p. 527]. A specific example of this analogy would be between a dead language embodied in a script which has not yet been deciphered and tools whose function is broadly recognised but for which the human knowledge of their specific use has not survived (and, historically, this has been asserted of a cooper's tools). From this perspective, the knowledge embodied in technology is not located in the technology or its makers and users but must be continually recreated in the interaction between them. Knowledge, like the technologies in which specific forms of knowledge are embodied, can appear as a relation between things or as relation between men and things but is '*a social relation between men*' [22, p.54].

Forms of knowledge resistant to symbolic articulation have been significant to the construction and use of technologies [26, p. 78]. At one level, certain forms of technology, such as 19th century United States agriculture, have been resistant to book learning, although this could be connected more directly with the relative cognitive authority attributed to local speakers and to books [48] and with the locally variable conditions for agriculture [26, p. 226]. At a further level, the knowledge embodied in the development and use of technology may not lend itself to symbolic or theoretical articulation [40, p. 69]. Similarly, the 'theoretical conception [of how to execute a particular process] must be perfected by accumulated experience on a large scale' [22, p. 502]. The resistance of such knowledge to symbolic articulation may explain the significance of the migration of personnel in technology transfer [26, pp. 154-155], although patterns of emigration have also been significant to the development of scientific, symbolically articulated, knowledge. With the late 19th century convergence of science and technology [39, pp. 1–3], symbolically articulatable knowledge may have become more significant to advanced technology, and this would clearly be true of sophisticated forms of information technology, but the strong possibility of continuity with knowledge resistant to symbolic articulation must be allowed. Where knowledge is not reducible to symbolic form, which is then open to separation from its producer, it may retain its human origins more transparently.

An example of a technology in use, of a steamboat on the Mississippi, may help to understand and substantiate these themes:

> 'Did you ever know of a boat following a bend upstream at this stage of the river?'
>
> 'No, sir, – and I was n't trying to follow it. I was getting away from a bluff reef.'
>
> 'No, it was n't a bluff reef; there is n't one within three miles of where you were.'
>
> 'But I saw it. It was as bluff as that one yonder.'
>
> 'Just about. Run it over.'
>
> 'Do you give it as an order?'
>
> 'Yes. Run over it.'
>
> 'If I don't, I wish I may die.'
>
> 'All right. I am taking the responsibility.'
>
> I was just as anxious to kill the boat, now, as I had been to save her before. I impressed my orders upon my memory, to be used at the inquest, and made a straight break for the reef. As it disappeared under our bows I held my breath; but we slid over it like oil.'
>
> 'Now don't you see the difference? It was n't anything but a *wind* reef. The wind does that.'
>
> 'So I see. But it is exactly like a bluff reef. How am I ever going to tell them apart?'
>
> 'I can't tell you. It is an instinct. By and by you will just naturally know one from the other, but you will never be able to explain why or how you *know* them apart.'
>
> It turned out to be true. The face of the water, in time, became a wonderful book – a book that was a dead language to the uneducated passenger, but which told its mind to be without reserve, delivering its most cherished secrets as clearly as if uttered them with a voice [49, pp. 93–94].

It may be worth recalling that Wiener's deliberate reinvention of the term *cybernetics* was a testimony to the pilot's concern with control mechanisms [29, 50].

The non-verbal signals read from the river surface are made into signs by human interpretation. Those signs can be spoken about but cannot be given direct verbal equivalents – 'you will never be able to explain why or how you *know* them apart'. For passengers, the surface of the water remains a 'dead language', whose significant, rather than physical, aspect, may not be even be known to them, and who, even given this awareness, might not be able to isolate individual signs. An analogous situation existed historically with cuneiform, which was received as natural marks rather than a human semiotic construction. Without the semiotic interpretation, the boat remains inert machinery, loses both use- and exchange-value when deprived of contact with living labour [33, p. 527]. In terms of the previously established distinctions with regard to information technology, the concern here is with control mechanisms in a transport technology.

The knowledge involved in the piloting of the boat is transmitted by direct apprenticeship, whose completion requires an immersion in detailed experience. The design and construction of the boat itself embodies the cumulative experience of its constructors. In use, the boundary between the machine and its environment is a matter of convention, a theme present both in studies of technology and of automata: for Marx, anticipating the discussion of the location of knowledge, for large scale machinery, 'the skill resides not in the worker but in the machine and in the scientific combination of both as a whole in the factory' [22, p. 529n]; analogously, in automata studies, the question of where a particular machine ends and its environment begins can only be settled by convention or definition [15, p. 19]. Despite the hard materiality of technology, it resembles a language, where the point of view creates the object [45]. The knowledge in use of the technology resides in the collective of pilot and apprentice, boat, and river surface. In the intervals between human interaction, the 'system of machinery ... constitutes in itself a vast automaton ... driven by a self-acting prime mover' [22], and here there is a specific analogy with one sense of non-determinism in automata theory, where a series of automatic transformations continues until an ambiguous configuration is reached and the process halts until a human choice is made [21].

Invention, innovation, and diffusion

Within the modern understanding of technology as a human construction, a series of systematic distinctions, derived from Schumpeter, is made with regard to the dissemination and adoption of technologies. At the greatest level of granularity, conceptualisation, invention, innovation, and diffusion are differentiated, understood as follows [26, pp. 192–193]:

- Conceptualisation – the idea for a technology.
- Invention – establishment of technical feasibility.
- Innovation – the commercially feasible product.
- Diffusion – the subsequent adoption and spread of the innovation.

In less finely differentiated, and possibly more standard, accounts conceptualisation is not distinguished as a separate phase [51]. A debate exists as to the sharpness or rigidity of the distinctions, and, contrary to a rigid interpretation of the Schumpeterian framework, it is argued that minor movements in a technology can have cumulative significance: innovation is only the beginning of the diffusion process and an innovation may be substantially changed in character in diffusion [26, pp. 66-81; 52, 53]. It is also acknowledged that the relation between these stages may change historically, for instance, with the transformation of invention from a private into a corporate activity. It is tempting to apply a phrase derived from considerations of the relation between expression and idea in United States copyright, a continuum of similarity, but still to regard distinctions along the continuum as indicating analytically valuable, and substantively real, differentiations.

The need for these distinctions in information science can be sensed from their absence in studies of information technology, with conceptual confusion resulting, and from their presence in inchoate form, without reference to the clarifying framework [40]. For instance, Warner anticipates the distinctions of invention, but without the incisiveness made possible by direct use of these positions:

> A final, admittedly imprecise but still valuable, dimension for differentiating forms and other systems of signs is their synchronic social significance. ... The mere existence of a sign system does not guarantee its social significance. For instance, the development of copyright in the sense of intellectual property in the United Kingdom and the United States suggests that the social significance of computer programming to those political communities grew gradually, impinging first on the judiciary and then emerging in legislative deliberations in the late 1970s, some thirty years after the development of the stored program computer [21, p. 40].

In this passage, existence would correspond to invention and social significance to the continuum from innovation to diffusion. A comparative reading would suggest firmer adoption of distinctions in media studies [54]. Their adoption within information science should clarify its own treatment of information technologies and also make available a framework within which the historical perspective of information, and, in this instance, library, science can be used to explore the socio-technical dynamic associated with information technologies.

The significance and value of the

distinctions may be better understood when they are compared with analogous distinctions made with regard to scholarly communication, which are well established within information science. For instance, conceptualisation could be regarded as analogous to a dialogue with oneself, or, at a further stage, informal conversations with colleagues; invention compared to reduction to writing; innovation to publication, which itself need only be regarded as a binary event for certain purposes [55, p. 897]; and diffusion to adoption of a text or idea by scholarly communities. Not distinguishing innovation from diffusion for information technologies would be tantamount to compounding publication and communication. Further, for both publication and adoption, there are known indicators (bibliographic records for publication and citation data for adoption). For these, an important distinction is made between the indicator and the phenomenon, which might escape detection by the indicator or reveal itself in other ways. Similar methodological reservations will be placed on the indicators to be proposed here, for assessing innovation and diffusion of information technologies.

In summary, the series of distinction from conceptualisation to diffusion are to be regarded as analytically valuable, if not essential, distinctions, for information as well as productive technologies. Historians of industrialisation have been preoccupied with invention, possibly, as it may be easier methodologically to isolate, in the first instance, and have neglected diffusion. Yet the social and economic consequences of technological changes are a function of their rate of diffusion and not the date of their first use. From the perspective of world history, the critical social process requiring examination is diffusion [26, p. 173]. As part of the study of the diffusion, we should seek to understand how the superiority of a technology came into being rather than ascribing it to inherent qualities of the technology. A similar focus of attention on invention can be discerned in studies of information technologies (for instance, Ohlman [40]) and a comparable shift would be equally desirable.

Summary

Technology is to be regarded as a radical human construction, adopted under specific historical circumstances. It is viewed as a social relation between men, which may appear as a relation between men and things or between things and things, but this false appearance should be dissolved. Information technology was differentiated from productive technology, by its concern with the transformation of signals, not natural resources. A further distinction between use of information technologies in human communication and in machine control was proposed. The concept of knowledge, either as existing in human interaction or as embodied in technology, was not reified but neither was the materiality of the machine denied. Crucial distinctions, relating to the invention and adoption of technologies, were introduced. Further projects, part of an information view of history [10], can now be foreshadowed.

Information view of history

From an information view of history, a number of crucial questions can be identified:

- First, given that diffusion, not invention, is taken to be of primary significance, what indicators for diffusion can be supplied?
- Secondly, what was the critical period for innovation and diffusion of modern information technologies (preliminary evidence identifies the late 19th century United States)?
- And, thirdly, why did innovation and diffusion occur at that time and in that place?

Other, comparably, although not necessarily equally, significant questions can be asked and have begun to be addressed [46], but these are sufficient opening topics, which can be briefly outlined here.

Registering of innovation and diffusion in copyright

The history of copyright represents a rich and largely, although not entirely [34, 56, 57], unexploited resource for tracing the development and impact of information technologies. A focus on copyright will give a bias toward human to human, and public not private, communication, rather than control mechanisms (although some media impacting on copyright, such as musical or piano rolls, could be regarded both as control mechanisms and as part of human communication). This distinction is at least acknowledged, in contrast to an exclusive focus on control mechanisms [35], and an emphasis on communication and transport as a whole may be more convincing. Patent would reveal only invention, and, although the chronological gap between invention and diffusion testifies to the significance of the distinction (consider the typewriter, for instance), the focus here is to be on diffusion and, to a lesser extent, on innovation.

Copyright registration and legal deposit provide, in differing ways, indicators of the proliferation of non-book forms. Copyright registration in the United States was essential to claims for copyright, until the *Berne Convention Implementation Act 1988* allowed copyright to subsist, but not necessarily to be fully asserted, without prior formalities [34]. Both legislation and, with greater numerical detail, copyright registrations then provide an explicit record of the assimilation of non-book forms to copyright protection. Legal deposit in the political regions governed by the United Kingdom (the pedantry of the phrase is only justified by its accuracy as deposit was extended throughout the British Empire) has not been fully required for copyright since the late 19th century. A less numerically precise indicator of the non-book forms is, however, provided by the gap between claims for copyright as intellectual property and works received under legal deposit [56]. The level of impact for non-book forms given by these indicators can be placed on the continuum from innovation to diffusion. For both jurisdictions, imperfect and historically variable, collection and recording practices will have to be considered [56, 58].

Some control mechanisms and codes used in human to human communication (for instance, telegraphic codes) have impinged on copyright [57] and this can yield a further indicator of the development of information technologies. A non-technical approach would be to regard the costs of litigation and of legislation as deterring action unless the artifact in question has acquired a certain level of social significance. A complementary approach, informed by the distinctions developed here, would be to regard copyright litigation as evidence of innovation, for there must be at least two competing parties disputing ownership of intellectual property (and not simply an invention without commercial significance). Even denial of protection is indicative of innovation. Copyright legislation would then be taken as indicating diffusion, with the social costs of producing knowledge in exact, written form [26, 46] taken as discouraging legislation in response to relatively isolated innovations. Empirically, this distinction of innovation from diffusion can be supported by the pattern of technologies impinging on the judiciary, traditionally regarded as oriented towards precedent, and the subsequent correction of anomalies arising by the legislature [57], concerned with current and future public good.

The history of copyright can, then, provide highly specific indicators of information technology innovation and diffusion. If attention is primarily focused on the United States and the United Kingdom the geopolitical limitations of the study must be acknowledged, although a geopolitical, as well as chronological, parameter is simultaneously provided. The indicator adopted needs to be distinguished carefully from the phenomenon, which, it is recognised, may elude the indicator or emerge in other ways. For fuller understanding, evidence derived from copyright would have to be viewed in relation to other indicators and placed in the context of developments in control, communication, and transport as a whole.

Critical period for modern information technologies

Evidence from the history of copyright which could illuminate information technology innovation and diffusion has been partially accumulated and made public by Warner [56, 57], but for contrasting, although related, purposes. The chronology presented here can be read diachronically or synchronically (see Appendix). The approach will be to read each category diachronically but to comment on the synchronic patterns rendered visible.

The formal history of legal deposit for the United Kingdom can be read from United Kingdom legislation. For the *Copyright Act 1842*, single, inherited definition of a book encompassed a work for deposit and for intellectual property; for the *Copyright Act 1911*, a book for legal deposit was differentiated from literary and other works for intellectual property. Although not directly revealed by the formal history of legislation, deposit began to be rigorously and extensively enforced from the 1850s and 1860s [56]. Closely concurrent, and possibly slightly earlier, signs of the diffusion of non-book forms can be found in United States legislation for media to be covered by copyright registration. The *Copyright Act 1870* conferred the registration of copyrights and the deposit of works on the Library of Congress, establishing registration procedures appropriate to the material form of the work. The divergence of legal deposit and copyright as intellectual property for the United Kingdom is confirmed by the 1911 Act; copyright registrations for the United States indicate the diffusion of non-book materials from the mid-1860s

Copyright cases in the United Kingdom are indicative of innovations in shorthand and telegraphic codes in the mid-1880s and of musical (or piano) rolls from the late 1890s. Despite their extensive diffusion in the United States, telegraphic codes do not emerge as subjects of judicial dispute until 1921. Musical or piano rolls feature in a well-known, if not notorious, case of 1908, *White-Smith v. Apollo*. In the case of musical or piano rolls, anomalies arising from the judicial denial of protection were deliberately corrected by legislation in 1909 (United States) and 1911 (United Kingdom) (a view that *White-Smith v. Apollo* was not fully overruled by the *Copyright Act 1909* did, however, dominate [57]). The *Copyright Act 1870* expanded the range of forms protected in the United States and from the 1870s onwards a liberal, or inclusive, interpretation was given to the constitutional clause, the 'writings of an author'. International conventions, indicative of increasing international communication begin from the 1880s. The relative autonomy of the United States is revealed in its refusal to accede to those conventions, although the possibility of recognising international copyright was opened up from 1891 [34]. Evidence from litigation (corresponding to innovation) and, more fully and convincingly, from legislation (indicative of diffusion) strongly suggest the 1870s as a crucial period for the development of modern information technologies, again with developments occurring slightly, but significantly, earlier in the United States.

Other developments, including those in the more narrowly construed information and library sphere, support the significance of the mid- to late 19th century. Panizzi's historical moment occurs when transport and communication had made the wide enforcement of legal deposit possible, although public communication was shortly to be driven beyond the forms captured by deposit. The Library of Congress was broadly concurrently transformed from a government library into an institution of national significance. H.G. Wells' speculation on the world brain [59] follow the empirical possibility of global, or near global, communication through signal transmission. These developments could be regarded as further instances of being, partly mediated through technology, determining consciousness. Information technology developments (derived from standard accounts of invention, which acknowledge that dates for inventions must be approximations) cluster from the mid-19th century, with a particularly intense phase in the 1870s. The presentation here, if read synchronically from left to right, does at least give ontological priority to forms of the social over the technical.

The 1870s and 1880s have, then, been compellingly identified as the crucial period for innovation and diffusion for modern information technologies, with the United States, and, to some extent, links between the United States and Europe, as highly significant loci. This is in sharp contrast with assumptions of the recency of modern information technology, which has influenced discussions of the information society concept [60]. However, it is congruent with other accounts: with Beniger's chronological, and geopolitical, location of the control revolution and account of the development of modern communication technologies [35, pp. 7–20]; with the understanding of globalisation as a late 19th century movement in communication studies [61]; and with Wells' early understanding of global communication [62]. The geopolitical, as distinguished from the chronological, locus identified seems to have been less anticipated.

Why at that time and in those places?

Radical changes in information and communication technologies, as distinct from cumulative developments in existing innovations, are rare in world history. Such technologies have more often constituted a relatively stable backdrop to more dynamic events [63, p. 63]. The question of why this nexus of developments should have occurred at that time and in those places must, then, be addressed.

The late 19th century United States is familiarly regarded as the age of invention. Mark Twain's personal involvement in unsuccessful innovative enterprises and fictional representations, in *The Gilded Age* and *A Connecticut Yankee in the Court of King Arthur*, are well known. For one aspect of the innovation of the period, the American system of manufacture, or the production of standardised goods from specialised machines with interchangeable parts, an explanation has been given for its emergence first in the United States. On the demand side, high aggregate demand for certain classes of commodity, which could be connected with the expanding population of the period, enabled production processes whose profitability required long runs. The limited variety in composition of the demand encouraged the use of highly specialised machinery. On the supply side, there was a rich abundance of natural resources, not reduced by extensive consumption by previous technological developments. Processes which were resource wasteful, but economically rational, could then be adopted [28, pp. 109-120]. Readers of Fenimore Cooper will recall reiterated references to 'the white man and his wasty ways' and the conditions of the period have been suggested as the roots of American profligacy with natural resources. In summary, labour scarcity and natural abundance pushed rapidly towards rich inventive possibilities [28, 109-120].

Further conditions for technological dynamism and creativity can be distinguished. The essential dynamic of capitalism is innovatory:

> Modern industry never views or treats the existing form of a production process as the definitive one. Its technical basis is therefore revolutionary, whereas all earlier modes of production were essentially conservative [22, p. 617].

Capitalism, in this sense, emerged in particularly pure form in the late 19th century United States:

> Nowhere does the fluidity of capital, the versatility of labour and the indifference of the worker to the content of his work appear more vividly than in the United States of North America. In Europe, even in England, capitalist production is still affected and distorted by hangovers from feudalism [22, p. 1014n].

A contrast is made with the merely stable labour power of the slave. The account of labour mobility and the fluidity of capitalism is consonant with the fictional representations of the period. Crucial dimensions would be expanding territory and population and the absence of a constraining inheritance.

These conditions would enable invention and innovation in information and communication technologies, but not, given the high rate of failure in innovation, account for their diffusion and adoption. The pivotal position of the market, and of communication in the market (commerce and communication share a common etymology), must be recognised [24]. A further, more historically specific, demand for signal transmission over space would have been created by frontier expansion and the developing links with Europe. The rapid diffusion of the telephone and the telegraphy can, in different ways [64], be understood in this context. The closing of the frontier is dated to 1880, shortly after the diffusion of a nexus of communication technologies and concurrent with their continuing development. A thematic link is thereby created to the, still powerful, frontier hypothesis, which takes the existence of an area of free land (assimilating, in its original form, conquest to expansion) and the continual recession of the frontier as crucial to American economic and cultural development [34, 65, 66].

A preliminary explanation has, then, been given for the diffusion of information and communication technologies in the late 19th century United States and between the United States and north-west Europe. Even if other geopolitical regions were to be identified as sites for innovation and adoption (dates for diffusion in Europe, implied by the proliferation of international treaties from the 1880s, indicate a slight, but still significant, lag), the explanation for the United States would remain valid. The converse or absence of the crucial factors identified might be relevant to the suggested lag, for instance, the legacy of feudalism in Britain [67], and for France, with its large internal distances, the envisioning of communication as proceeding from the state to the individual [61, 68], in contrast to American democracy and populism.

Conclusion

A largely novel view of information technology as a radical human construction, carefully differentiated from productive technology, has been developed, drawing on themes derived from Marx, and mediated by economics, and combining these with understandings established in automata studies and information science. An account was given of critical distinctions, from invention to diffusion, developed in economics, which could then be applied to information technologies. Data from the history of copyright, and, for the United Kingdom, from legal deposit, little known outside information and library science, was used compellingly to suggest the late 19th century United States as a crucial time and place for modern information technology innovation and diffusion An explanation for the adoption of information technologies was given in terms of the technological dynamism and geopolitical expansion and communication of the period.

The understanding of information technology developed needs refinement, and the evidence and explanation given for its diffusion requires expansion and development, but the possibility of a valuable intra- and extra-disciplinary contribution has been opened up.

References

1. Roberts, N Social considerations towards a definition of information science. *Journal of documentation*, 32(4), 1976, 249-257.

2. Rosenberg, V. The scientific premises of information science *Journal of the American Society for Information Science*, 25(4), 1974, 263–269.

3. Brown, A.D. *Towards a theoretical information science: information science and the concept of a paradigm*. Sheffield: Department of Information Studies, University of Sheffield, 1987.

4. Levinson, P. *The soft edge: a natural history and future of the information revolution*. London and New York: Routledge, 1997.

5. Warner, J. Review of Paul Levinson. The soft edge: a natural history and future of the information revolution. *Journal of the American Society for Information Science*, 49(1), 1998, 1041.

6. Hancock-Beaulieu, M., ed. *Information systems for end-users: research and development issues*. London and Los Angeles: Taylor Graham, 1992

7. Warner, J. Review of M. Hancock-Beaulieu. Information systems for end-users: research and development issues. *Journal of documentation*, 49(4), 1993, 421–424.

8. Kahin, B. and Nesson, C., eds. *Borders in cyberspace: information policy and the global information infrastructure*. Cambridge, MA: The MIT Press, 1997.

9. Warner, J. Review of B. Kahin and C. Nesson, eds. Borders in cyberspace: information policy and the global information infrastructure. *Journal of the American Society for Information Science*, 49(4), 1998, 387–388

10. Warner, J. An information view of history. *Journal of the American Society for Information Science*, 50(12), 1999, 1125-1126.

11. Wyn, E. and Katz, J.L. Hyperbole over cyberspace: self-presentation and social boundaries in Internet home pages and discourse. *Information Society*, 13(4), 1997, 297–327.

12. Gilfillan, S.C. *The sociology of invention.* Cambridge, Massachusetts: MIT Press, 1970.

13. Winner, L. *Autonomous technology: technics-out-of-control as a theme in political thought.* Cambridge, Mass: MIT Press, 1977.

14. Walsham, G. IT and changing professional identity: micro-studies and macro-theory. *Journal of the American Society for Information Science*, 49(12), 1998, 1081–1089.

15. Minsky, M. *Computation: finite and infinite machines.* Englewood Cliffs, NJ: Prentice-Hall, 1967.

16. Orwell, G. *Charles Dickens.* In. Orwell, G. *Dickens, Dali & others.* New York: Harcourt Brace & World, Inc., 1946. 1–75.

17. Wiener, N. *Invention: the care and feeding of ideas.* Cambridge, Massachusetts and London, England: MIT Press, 1993.

18 Warner, J. W(h)ither information science?/! *Library Quarterly.* To appear 2001.

19. Dijkstra, E.W. On the cruelty of really teaching computer science (The SIGCSE Award Lecture 1989). *SIGCSE Bulletin: a Quarterly Publication of the Special Interest Group on Computer Science Education*, 21, 1989, 1, xxiv-xxxix.

20. Leith, P. *Formalism in AI and computer science.* New York: Ellis Horwood, 1990.

21. Warner, J. *From writing to computers.* London and New York: Routledge, 1994.

22. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Ernest Mandel. Translated by Ben Fowkes. Harmondsworth: Penguin Books in association with New Left Review, 1976.

23. Buhle, P. *Marxism in the United States: remapping the history of the American Left.* Second edition revised. London and New York: Verso, 1991

24. de la Haye, Y., ed.. *Marx & Engels on the means of communication (the movement of commodities, people, information & capital): a selection of texts.* New York: International General and Bagnolet, France: International Mass Media Research Centre, 1980.

25. Mackenzie, D. *Knowing machines: essays on technical change.* Cambridge, Massachusetts and London, England. MIT Press, 1996.

26. Rosenberg, N. *Perspectives on technology.* Cambridge: Cambridge University Press, 1976.

27 Rosenberg, N. *Inside the black box: technology and economics.* Cambridge: Cambridge University Press, 1982.

28. Rosenberg, N. *Exploring the black box: technology, economics and history.* Cambridge: Cambridge University Press, 1994.

29. Wiener, N. *The human use of human beings: cybernetics and society.* New York: De Capo Press, 1954.

30. Kling, R., Rosenbaum, H. and Hert, C. Social informatics in information science: an introduction. *Journal of the American Society for Information Science.* 49(12), 1998, 1047–1052.

31. Stewart, J. and Williams, R.. The coeveolution of society and multimedia technology: issues in predicting the future innovation and use of a ubiquitous technology. *Social Science Computer Review.* 16(3), 1998, 268–282.

32. Marx, K. *Grundrisse: Foundations of the critique of political economy (Rough draft).* Translated with a Foreword by Martin Nicolaus. London: Penguin Books in association with New Left Review, 1973.

33. Marx, K. *Capital: a critique of political economy.* Volume Three. Introduced by Ernest Mandel. Translated by David Fernbach. Harmondsworth: Penguin Books in association with New Left Review, 1981.

34. Warner, J. Information society or cash nexus? A study of the United States as a copyright haven. *Journal of the American Society for Information Science.* 50(5), 1999, 461–470.

35. Beniger, J.R. *The control revolution: technological and economic origins of the information society.* Cambridge, Massachusetts and London, England: Harvard University Press, 1986.

36. Lukács, G. Technology and social relations. In: *Marxism and human liberation: essays on history, culture and revolution.* Edited with an introduction by E. San Juan, Jr. New York: Delta, 1973, 49-60.

37. Marx, K. The eighteenth Brumaire of Louis Bonaparte. In: Mark, K. *Surveys from exile: political writings.* Edited by David Fernbach. Harmondsworth: Penguin Books in association with New Left Review. Volume 2, 1973, 143–249.

38. Vico, G. *On the most ancient wisdom of the Italians: unearthed from the origins of the Latin language: including the disputation with the Giornale de' Letterati d'Italia.* Translated and introduced by L.M. Palmer. Ithaca and London: Cornell University Press, 1988.

39. Mowery, D.C. and Rosenberg, N. *Paths of innovation: technological change in 20th century America.* Cambridge: Cambridge University Press, 1998.

40. Ohlman, H. Information: timekeeping, computing, telecommunications and audiovisual technologies. In: McNeil, I., ed. *An encyclopaedia of the history of technology*. London and New York: Routledge, 1996, 686-758

41. Shannon, C E. and Weaver, W *The mathematical theory of communication*. University of Illinois Press: Urbana, 1949.

42. Bar-Hillel, Y. An examination of information theory In. Bar-Hillel, Y. *Language and information. selected essays on their theory and application*. Reading, Massachusetts: Addison-Wesley and Jerusalem: Jerusalem Academic Press, 1964.

43. Buckland, M. Information as thing. *Journal of the American Society for Information Science*. 42(5), 1991, 351–360.

44. Warner, J. The public reception of the Research Assessment Exercise 1996. *Aslib Proceedings*. 49(10), 1997, 263–276.

45. de Saussure, F *Course in general linguistics*. Edited by C. Bally and A. Sechehaye with the collaboration of A. Riedlinger, translated and annotated by R. Harris. London: Duckworth, 1983.

46. Warner, J. Not the exact words· writing, computing, and exactness. In: Warner, J *Information, knowledge, text*. Metuchen, NJ Scarecrow Press, 2000, 33–46 (To be published in November 2000.)

47. Collins, R. *The sociology of philosophies: a global theory of intellectual change* Cambridge, Massachusetts and London, England. The Belknap Press of Harvard University Press, 1998.

48. Wilson, P. *Second-hand knowledge: an enquiry into cognitive authority* (Contributions in Librarianship and Information Science 44). London and Connecticut: Greenwood Press, 1983.

49. Twain, M. *Life on the Mississippi* New York. Viking Penguin, 1984.

50. Warner, J. Shannon, Weaver, and the Minotaur In: Warner, J. *Information, knowledge, text*. Metuchen, NJ: Scarecrow Press, 2000, 17–31. (To be published in November 2000.)

51. Freeman, C. Innovation. In: Eatwell, J., Milgate, M. and Newman, P., eds *The new Palgrave: a dictionary of economics*. London and Basingstoke· Macmillan, 1987. Published in the United States of America and Canada by the Stockton Press, New York. 4 Volumes. Volume 2, 858–860.

52. Nelson, R.R. Innovation. In: Sills. D.L., ed *International encyclopedia of the social sciences*. New York: Macmillan Company and The Free Press, 1969 Volume 7, 339-345.

53. Stolpher, W.F. and Schumpter, Joseph A. In: Sills, D.L., ed. *International encyclopedia of the social sciences*. New York: Macmillan Company and The Free Press, 1969. Volume 14, 67–72.

54. Howells, R. A movie murder mystery. *The Times Higher*. 23 July 1999, 18–19.

55. Kling, R. and McKim, G. Scholarly communication and the continuum of electronic publishing *Journal of the American Society for Information Science* 50(10), 1999, 890–906.

56 Warner, J. Writing, literary work and document in United Kingdom copyright. *Journal of Information Science*, 16(5), 1990, 279–289.

57. Warner, J. Writing and literary work in copyright: a binational and historical analysis. *Journal of the American Society for Information Science*. 44(6), 1993, 307–321.

58. Solberg, T *Copyright miscellany. being the later contributions concerning the protection of literary and artistic property with autobiographical sketch, portrait and bibliography*. Boston: John W. Luce & Company, 1939.

59. Rayward, W B. H.G Wells's idea of a World Brain: a critical reassessment. *Journal of the American Society for Information Science* 50(7), 1999, 557–573.

60. Webster, F *Theories of the information society* London and New York: Routledge, 1995.

61. Mattelart, A *The invention of communication*. Translated by Susan Emanuel. Minneapolis, MN: University of Minnesota Press, 1996.

62 Wells, H.G. *Anticipations of the reaction of mechanical and scientific progress on human life and thought*. London: Chapman and Hall, 1902.

63. Collins, R *The credentials society: a historical sociology of education and stratification*. New York: Academic Press, 1979.

64. de Sola Pool. I , ed *The social impact of the telephone*. London: England and Cambridge, Mass.: MIT Press, 1977.

65. Turner, F.J. The significance of the frontier in American history. In: Turner, F.J. *The frontier in American history*. New York: Dover Publications, 1996, 1–38.

66. Grossman, J.R., ed. *The frontier in American culture*. (An exhibition at the Newberry Library, August 26, 1994 – January 7, 1995). Berkeley, Los Angeles and London: University of California Press, 1994

67. Porty, C.R. The British experience 1876–1912: the impact of the telephone during the years of delay. In: de Sola Pool, I., ed. *The social impact of the telephone*. London: England and Cambridge, Mass.: MIT Press, 1977, 69–96.

68. Attali, J. and Stourdze, Y. The birth of the telephone and economic crisis: the slow death of monologue in French society. In: de Sola Pool, I., ed. *The social impact of the telephone*. London: England and Cambridge, Mass.: MIT Press, 1977, 97–111.

Appendix: chronology for innovation and diffusion for modern information technologies

Date	United States legislation	United Kingdom case	United Kingdom legislation	International convention	Other developments	Information technology development
1710			Copyright Act (1709)			
1789	Constitutional provision concerning copyright					
1790	Copyright Act (1790) The Act granted 'author and authors of any map, chart, book or books .. the sole right and liberty of printing, reprinting, publishing and vending such map, chart, book or books' (Copyright Act, 1790).					
1826						Photography
1831	Copyright Act (1831).					
1839						Micro-photography
1842			Copyright Act (1842) A work in which copyright was to be 'construed to mean and include every volume, part or division of a volume, pamphlet, sheet of letterpress, sheet of music, map, chart, or plan separately published'.			
1844						Baltimore – Washington telegraph
1846					Establishment of Smithsonian Institution.	
1854						Sounder telegraph
1856					Panizzi appointed Principal Librarian of the British Museum (held office 1856–1866).	
1860						Typewriter (reliable). Pony Express.
1861						Trans-continental telegraph
1865	Copyright Amendment (1865). Extended copyright to photographs and their negatives.			International Telegraph Union		
1866						Trans-Atlantic cable (successful)

Date	**United States legislation**	**United Kingdom case**	**United Kingdom legislation**	**International convention**	**Other developments**	**Information technology development**
1870	Copyright Act (1870). Extended protection to 'painting[s], drawing[s], chromo[s], statue[s], statuary, and ... models or designs intended to be perfected as works of the fine arts' Library of Congress acquired registration of copyrights and the depositing of works.				Period of transformation of Library of Congress into an institution of national significance (had been created in 1800 to serve the information needs of government).	
1876						Telephone
1878						Phonograph
1880					Closing of United States frontier.	
1880						Hollerith punched card machine
1882	Copyright Amendment (1882).					
1884		Pitman v. Hines Concerned with copyright in a system of shorthand (referred to as phonograph). Ager v. Peninsula. Concerned with codes for telegraphic transmission				
1886				Berne Convention (1886)		
		Ager v. Collingridge. Concerned with codes for telegraphic transmission.				
1891	Copyright Amendment (1891). Non-residents and non-citizens, not entitled to claim copyright in the United States copyright since 1790, could be recognised as copyright holders, after bilateral agreements with their country of citizenship.					
1895						Cinematography
1896						Wireless telegraphy
1897						Pianola (player piano) patented in United States
1899		Boosey v. Whight. Concerned with musical (piano) rolls.				
1901						Radio
1908				Berlin Convention (1908). Copyright could subsist to include 'any production in the literary, scientific or artistic domain, whatever may be the mode or form of its reproduction ... dramatic or dramatic-musical works, choreographic works and entertainments in dumb show, the acting form of which is fixed in writing or otherwise' (Berlin Convention, 1908, Article 2).		

Date	United States legislation	United Kingdom case	United Kingdom legislation	International convention	Other developments	Information technology development
1909	Copyright Act 1909. A musical roll could constitute a copy of a score, the musical roll itself, and comparable contrivances were not to be regarded as writings.					
1911			Copyright Act (1911). The category of a 'literary work' was introduced to United Kingdom copyright legislation by the 1911 Act and replaced that of the book in 1842 Act, as an artefact in which copyright could subsist. A literary work was to be interpreted as including 'maps, charts, plans, tables, and compilations' (Copyright Act, 1911, § 35 (1)). The definition of a book for deposit was largely unaltered from the 1842 Act but confined to a separate section. Records, perforated rolls, and other contrivances by means of which sound may be mechanically reproduced' were to be regarded as musical works (Copyright Act, 1911, § 19 (1)).			
1917		Anderson v. Lieber. Concerned with codes for telegraphic transmission.				
1928					Rome Convention (1928)	
1944						Programmable computer
1948					Brussels Convention (1948).	
1952					Universal Copyright Convention (1952).	
1956			Copyright Act (1956). Writing was to be interpreted as 'any form of notation, whether by hand or by printing, typewriting or any similar process.'			

Date	United States legislation	United Kingdom case	United Kingdom legislation	International convention	Other developments	Information technology development
1971				Paris Convention (1971)		
1976	Copyright Act (1976).					
1980	Copyright Amendment 9180).					
1985	Berne Convention Implementation Act (1988).		Copyright (Computer Software) Amendment Act (1985).			
1988			Copyright, Designs and Patents Act (1988).			

Subscriptions
Aslib Proceedings is published ten times per annum and supplied to corporate members of Aslib. For 2000 the annual subscription rates are £170/$265 (£136/$212 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 20 7903 0000. Fax: +44 (0) 20 7903 0011. Email: aslib@aslib.com *http://www.aslib.com/*

Periodicals Postage Paid at Rahway NJ and at additional mailing offices
POSTMASTER Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

© 2000 Aslib and contributors ISSN 0001 253X

0001-253X (2000) 52:9 ;313-370

ISSN 0001 253X
Volume 52
Number 10
November/December 2000

Cend Ans 12/04/01

Aslib Proceedings

New information perspectives

An evaluation of the impact of the Internet in Africa
B.M. Chivhanga

The automation of controlled vocabulary subject indexing of medical journal articles
David Roberts and Clive Souter

Narrative knowledge: eliciting organisational knowledge from storytelling
Stuart Hannabuss

A future for publishers serving technology industry markets: adapting to the opportunity of the Internet era
Paul M. Evans

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

Aslib Proceedings is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In some cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000–8,000 words in length. Articles should be accompanied by a 100–150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The editor is happy to provide advice on an informal basis to potential authors.

Editorial Board

Professor David Nicholas, *Head of Department, Department of Information Science, City University*; John Akeroyd, Director, *Learning Resources, South Bank University*; Sarah Blair, *Head of Publications, Aslib*; Peter Chapman, *Information Manager, NewQuest North East*; Professor Peter Cole, *Head of Department, Department of Journalism, University of Sheffield*; Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw*; Harry East, *Senior Research Fellow, Department of Information Science, City University*; Mike Hyman, *Managing Director, Head Software International*; Robin Hunt, *Creative Director, AreHause*; Nat Lievesley, *Network Manager, Centre for Policy on Ageing*; Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Senior Lecturer, Department of Information Science, City University*; Richard Withey, *Director, Independent Digital*; Ralph Adam, *Freelance Journalist;* Professor David Ellis, *Department of Information and Library Studies, University of Wales;* Professor Ben Fouche, *Stellenbosch University.*

Editorial Team

Editor: David Nicholas
Deputy Editor: Ian Rowlands
Editorial Manager: Kay Nicholas
Editorial Assistants: Peter Williams (City University), Paul Huntington (City University)

CONTENTS

373–383 An evaluation of the impact of the Internet in Africa
B.M. Chivhanga

384–401 The automation of controlled vocabulary subject indexing of medical journal articles
David Roberts and Clive Souter

402–413 Narrative knowledge: eliciting organisational knowledge from storytelling
Stuart Hannabuss

414–421 A future for publishers serving technology industry markets: adapting to the opportunity of the Internet era
Paul M. Evans

An evaluation of the impact of the Internet in Africa

B.M. Chivhanga
The Internet Studies Research Group (ISRG), Department of Information Science, City University, Northampton Square, London EC1V OHB
bmmc@soi.city.ac.uk

The article describes the diffusion of the Internet in Africa and its impact on its people, organisations and socio-economic activities. It shows which countries are leading the way and why, and more importantly the problems confronted by people in different countries. A conceptual framework based on the work of Press, Daly and Menou is used to analyse the diffusion of Internet technologies in Africa. A case study, involving the application of web technologies in the agricultural sector, is used as an example of the key role the Internet can play in respect to key socio-economic activities in Africa. The article concludes by looking at the future developments of the technology.

Introduction

The last few years have seen a phenomenal growth of the Internet in Africa. It has already lead to a transformation in the lives of large numbers of people and is fast changing the way organisations communicate and do business. What is however noticeable, when one looks at what is actually happening on the ground, is that there is now a full awareness of what the Internet is and what it can do. Many people are aware of the changes it has rung outside Africa. Of particular note is the fact that large organisations (both in the public and private sectors) now have an information policy that centres on the Internet as the main vehicle for communication and the dissemination of information within the organisation and with other organisations. Internet cafes have made an appealing presence as the latest thing in town. What is very noticeable is that there are distinct and varied groups of Internet users in each country. There are also marked differences in the rate of the up-take of the technology between individual African countries and also between the regions – North, South, East and West. It is now an accepted fact in Africa that the Internet changes 'everything'. There are a number of stumbling blocks preventing the realisation of the many hoped for Internet dreams and at times the Net experience is a very frustrating one as access speeds are notoriously slow.

Basis of the analysis

The general overview of the Internet in Africa presented here is based on Press's [1] analytical framework of the diffusion of Internet technologies in a given environment, Daly's [2] and Menou's [3] impact assessment model of new technologies in developing settings.

Press's framework comprises six categories namely: pervasiveness, geographic dispersion, sectoral absorption, connectivity, organisational infrastructure and sophistication of use. *Pervasiveness* relates to the level of use by non-technicians and the users' per capita income. The *geographical dispersion* looks at the concentration of the technology within a nation, analysing uses in cities and points of presence in the country. *Sectoral absorption* is concerned with the use of the Internet in education, commerce, health and the public sector, i.e. key development areas as identified by the United Nations Human Development Index. *Connectivity* infrastructure is based on the speed of Internet traffic and bandwidth to national and international backbones. *Organisational infrastructure* is all about the ISP industry and the conditions in the market. The last category, *sophistication*, evaluates the sophistication of use and how people have moved from being novice users to experienced users innovating as a result of their changing demands. This characteristic ties in very well with Menou's impact assessment model that looks at the changes in the use of new technologies over time – from being a novel experience to something that brings about lasting transformations in the behaviour of people.

Daly extends the analysis of Internet diffusion by looking specifically at unique conditions found in developing countries that call for different approaches. According to him, the approaches used to analyse the impact of the Internet in developing countries needs to take in the income and literacy levels of people and how the technology is being used to reduce poverty. His impact analysis model takes into account savings on telephone calls and faxes, changes in individual business performances, government agencies, schools, impacts in different sectors and markets, progress in the economy, health services, knowledge acquisition and the environment. He looks at both non-technological and technological benefits of using the Internet as a measure of real impact. Both Daly's and Menou's impact assessment conceptual frameworks look at changes resulting from Internet use on individuals, growing to include groups of people in one locality and spreading to large communities and eventually the rest of society.

Building upon the three models a unified impact assessment framework can be developed to analyse the diffusion of the Internet in developing countries, especially in Africa. This framework has four dimensions to it. Dimension 1 measures the prevailing pre-Internet connectivity infrastructure and the enabling environment existing in a particular country and region. Dimension 2 looks at the Internet connectivity industry, i.e. ISPs and the current market conditions. Dimension 3 concentrates on the general utilisation of the Internet in different socio-economic sectors as identified by Press and behavioural transformations over time as exemplified by Menou. In this dimension the national readiness to harness the technology, and non-technological changes resulting from the introduction of the Internet, will also be considered. Last, but not least, is Dimension 4, which focuses on the global developments of ICTs and their impact on National Information and Communications Infrastructure (NICI).

Figure 1 *A Conceptual framework to analyse the diffusion of the Internet in Africa*

Dimension	**Description**
Dimension 1	Pre-Internet connectivity infrastructure
Dimension 2	ISP industry and market conditions
Dimension 3	General utilisation and sectoral applications
Dimension 4	Global developments of ICTs

Pre-Internet connectivity issues include the inherited telecommunication legacies and the regulatory environment and how they impact on the ISP industry and the consumer market. The general utilisation of the Internet is concerned with how the key socio-economic sectors are using the new technologies and the benefits realised. It also covers:

- issues dealing with access to the Internet, the quality and quantity of web content and how this is changing the information environment;
- the national readiness in the uptake of the new technology, which examines the socio-economic circumstances of the people and the rate of adaptation to novel processes;
- people's income levels, per capita income, educational attainments, literacy levels and their general cultural characteristics and responsiveness to changes;
- the non-technological impacts of the Internet – the direct and indirect effect of the Internet on economic growth and reduction in poverty, how the Internet is contributing to wealth creation through ecommerce, for example, and a general improvement in the quality of living as a result of the application of the Internet in core traditional socio-economic activities.

Global developments of ICTs have a direct bearing on the development of national information infrastructures and this is crucial in Africa where technological innovations tend to come from outside the continent.

The uptake of the Internet in Africa – Dimension 1

Up until the end of April 1999, only three countries (Congo, Eritrea and Somalia) did not have local access to the Internet. These countries have since gained local Internet access. Political destabilisation and civil wars in these countries has virtually destroyed all telecommunication infrastructures, making it next to impossible to have full Internet connectivity in the parts of the country outside the capital cities. In fact throughout much of Africa Internet access is largely found in the capital cities. Exceptions do occur. Most notably in Zimbabwe, Tanzania, Mozambique, Kenya, Ghana, Botswana, Angola and Benin where Internet access has been extended to some secondary towns. South Africa, as might be expected, has Points of Presence (PoPs) in over 70 locations [4].

The number of Internet users continues to increase throughout Africa, even though at a much slower rate than in the world at large. The number of dial-up accounts readily available from ISPs shows that there are well over half a million regular users. The real figure is much higher than that. On a recent field trip to Zambia the author noted that an average dial-up account services three to four family members of the immediate family, as well as relatives and teenage friends. The actual number of Internet users for that one single Internet account could be up to five to ten people and this actually brings the total number of Internet users to about two to five million. This is just the author's estimate. However, a study conducted by the UN Economic Commission for Africa (UNECA) estimates that each Internet account supports up to three users and this gives a total of about 1.2 million Internet users in the whole of Africa. This figure must be considered a gross misrepresentation as ISPs and Internet cafes mushroom in many African capitals and as the young local educated population increasingly find that having an Internet account and using the Internet for getting information is more efficient and less time consuming than obtaining it from the traditional media.

There are significant country and regional differences in the number of Internet users. Countries such as Egypt have over 100,000 subscribers [5], Morocco – over 20,000, Kenya – over 15,000, Ghana and Zimbabwe – over 10,000. These countries have some of the highest numbers of users when compared to the rest of the African countries – the numbers are still very low when compared to the rest of the world. Africa is in danger of being left behind as the technological revolution gains pace. Northern and Southern Africa have the largest concentrations of users. The stronger the economy and the better the infrastructure, the greater the density of Internet users.

Another key reason why the Internet has not taken off in Africa as in other world regions, even though the local people are now keenly aware of its potential impact, is affordability. Telephone costs are extremely high in many African countries. A local call costs about US$0.17 a minute in Zambia, $0.48 in Angola, $1.30 in Benin and $0.40 in Cameroon. Where local call rates are lower one, unsurprisingly, finds higher populations of Internet users. For example, in Algeria a local call costs $0.02 a minute, in Botswana $0.06 and in Egypt $0.01 [5].

A major cost that is proving to be a stumbling block is the exorbitant charges of local ISPs. Local dial-up accounts can cost as much as $60 per month. That is a lot of money when converted to the local currency in many African countries and only the rich can afford it. Owning a computer in itself is a luxury and only the well off can afford one. In consequence the route to widening Internet access lies through the provision of cybercafes, kiosks and telecentres. A new feature in Southern African countries, such as Zambia, Botswana and Zimbabwe, is the installation of PCs with Internet access in post offices, hotels, lodges and clinics. Customers pay so much to send and receive email and surf the Web. These facilities are also found in many shopping malls and they are always in demand. These alternative ways to access the Internet in Africa have gone unaccounted for by organisations such as Network Wizards that only count the number of Internet users based on dial-up accounts. The total number of Internet users could be well over ten million throughout Africa, including South Africa.

Countries that have introduced telecommunication policies geared towards liberalising the sector have also enjoyed a relative reduction in the cost of accessing the Internet. Most African countries are entering into partnership with the private sector to improve the telecommunication infrastructure. This will eventually increase the total international bandwidth with countries such as Egypt, Kenya, Mauritius, Namibia, Senegal, South Africa and Tanzania seeing their traffic moving at 1Mbps. The total international bandwidth in Africa is about 34Mbps. Many people have to share that bandwidth and this slows down the connections.

Telecommunication developments that are underway to improve the speed of accessing Internet services include VSAT (very small aperture terminals) that will provide direct links to Europe or the US. Such countries as Zambia, Uganda, Tanzania, Ghana and Mozambique are developing positive regulatory environments, paving the way for the deployment of innovative telecommunications technologies geared to speeding up Internet traffic.

The ISPs industry and market conditions – Dimension 2

Internet Service Providers are mushrooming throughout Africa's capitals. There were well over 300 ISPs providing public Internet subscriptions at the beginning of 1999. Countries such as Egypt, Kenya, Morocco, Nigeria, South Africa, Tanzania and Zimbabwe have 10 or more Internet Service Providers and about 20 countries have only one ISP operating. Only Ethiopia and Mauritius have a national ISP monopoly, as private companies are not allowed to resell Internet services. Other countries in the Sahel Region, such as Tunisia, Mali and Mauritania, have a number of public sector ISPs.

Throughout Africa there is a growing trend for PTOs (Public trading organisations) to move into the Internet market. In many Francophone countries the PTO operates on a joint venture with France Cable and Radio called Telecom-Plus. In countries where the PTO has established the international Internet backbone, it is generally the only operator. Exceptions do occur in the case of Zambia, South Africa, Mozambique and Cote d'Ivoire and Nigeria. Multinationals that have made their influence felt include Africa Online (http://www.africaonline.com), which has operations in Tanzania, Uganda, Zimbabwe, Ghana, Kenya and Cote d'Ivoire. Other multinationals like UUNET, which has branches in South Africa, Swaziland, Zimbabawe and Namibia, and Swift Global, which operates in Tanzania and Uganda, are all trailing behind Africa Online.

In Tunisia, for example, ATI provides Internet services for government offices and institutions and government agencies (http://www.ati.tn). IRSIT (Regional Institute for Computer Sciences and Telecommunications) provides Internet services to research centres (http://www.irsit.mrt.tn). CCK (Khawarizmi Computing Centre) caters for universities and institutions of higher learning. INBMI (National Institute of Office and Computer Technology) provides Internet services to primary and high schools. Nearly all public sectors have their own ISPs in Tunisia. There are around two commercial ISPs – Planet Tunisie (http://www.planet.tn), and 3S Global Net (http://www.gnet.tn).

The Tunisian model is probably worth emulating in other African countries, as such a pattern of providing Internet access reaches out to a majority of people and institutions in key public sector areas, such as health and education, who would not normally afford the services provided by commercial companies. The government's involvement in the ISP market in Tunisia is a realisation of the advantages offered by new information technologies. The country's President Ben Ali announced in November 1997 strategic plans to connect all schools and public libraries to the Internet. This is already happening in the country!

The strategy for Internet connectivity in

Tunisia, just like in most African countries, takes a three pronged approach: developing and modernising the infrastructure, establishing suitable mechanisms for organisation and regulation and capacity building that involves training, subcontracting and developing an in-depth knowledge about the technology. Most of the personnel running these African ISPs are local and that has to be good news for the future growth and sustainability of the industry.

Make it quicker – ISP Services

The slow access to Internet services is still a major constraint to the effective and efficient running of the technology. After all the Internet is all about immediacy. This is largely because of the high tariffs on international calls and the poor circuit capacity of the telephone lines. Around 17 countries have an international bandwidth of 512Kps or more. About 10 countries, namely, Egypt, Kenya, Mauritius, Morocco, Namibia, Nigeria, Senegal, South Africa, Tunisia and Tanzania, have their Internet traffic going out at 1Mbs or more.

Most African Internet sites are being hosted by servers located in the US or Europe. The cost of hosting websites locally is very high and subscribers are encouraged to take advantage of the free overseas hosting services. Other developments that are addressing the bandwidth problem involve the use of a DirectPC-type system that provides incoming bandwidth of 64Kps for about US$30-$1000 a month and this largely depends on usage. This type of service can deliver up to 8Mbps and the normal telephone circuit is used for all outgoing traffic. The equipment used to get these speeds uses a digital KU-Band or C-Band satellite television antenna that costs between $175 to $500, depending on the size required. A decoder card that is installed in the PC costs about $450. Such equipment is being used by the staff of overseas organisations operating in remote rural areas. They also use satellite mobile phones that are attached to their laptops, even when they are not on field trips. The equipment itself is expensive for the local African and it is hoped that with time this technology will be rolled out to deal with the problems associated with Internet speed.

In Southern Africa, the satellite equipment is provided by four local companies: GIT (http://www.co.za), Hixsat (http://www.hixsat.co.za), Siyanda (http://www.siyanda.co.za) and Infosat (http://www.infosat.co.za). Interpacket (http://www.interpacket.net) offers similar satellite services that cover larger regions in Africa. Local ISPs are using these low-cost TV satellite dishes to manage high volumes of Internet traffic and this reduces their operating costs and increases the access speed to the Web.

The suppliers of Internet infrastructure – those companies providing access to the international Internet backbone for local ISPs – are predominantly large international outfits and these include AT&T, BT, Global One/Sprint, UUNET/AlterNet, MCI, NSN, BBN, Teleglobe, Verio and France Telecom/FCR. PanamSat and Intelsat provide direct links to private and PTO groundstations in the US and the UK, instead of going via the local PTO.

Establishing regional Internet backbones would be one way to reduce dependency on Internet backbones located outside Africa. This would result in regional hubs being used by ISPs in different countries. Sharing of equipment costs would mean better services to customers. At the moment this is not possible because the international tariffs charged by local PTOs discourages ISPs to establish multiple international links. One sees the need for regional co-operation between PTOs in each region if the Internet is to grow and make a major impact in Africa.

Other developments that are afoot, even though still in their infancy, are ISDN and video conferencing services. Only South Africa, Egypt, Tunisia, Morocco and Seychelles boast such services. Voice over Internet (VOIP) services are not yet available and none of the telecom operators have voice over IP technology. Egypt directs its voice traffic to the US.

Spread the message: general applications – Dimension 3

Most ISPs are responding to customer demand by making available low cost email services. This is the most popular feature of the Internet in Africa. Most people use Hotmail, Yahoo or Excite. They are also expensive, as connection to the remote site has to be maintained. Those that can afford to have Internet access charge people for using their account to surf and use web based email.

The UN Economic Commission for Africa, which has been instrumental in accelerating the growth of the Internet, point out that there is about one incoming and one outgoing email per day of an average of three to four pages. These are mainly communications with people outside the continent. A recent survey by ECA indicates that 25% of email is replacing fax messages, while 10% are replacing phone calls and the other 65% represent new communications that would not have arisen without the email system.

NGOs, private companies and universities have the highest number of email users. The majority of users are nationals although their representation varies from country to country. In Zambia for example 44% were nationals, whereas in Ghana the figure was about 90%. Most users are male. In Ethiopia men account for about 86% of users, in Senegal the figure is about 87% and in Zambia about 64%. Just as in the most developed countries, the largest number of users are very educated. In Zambia and Ethiopia, about 87% and 98% respectively of the users were educated up to degree level. This pattern is reflected throughout Africa. In South Africa the average user is male – 26 to 30, speaks English, is a high school or university graduate and earns between US$24,000 to US$40,000 and is a computer professional.

Email is the facility that has made the greatest impact in Africa. It is being used both for personal and business purposes. It is becoming by far the best means of communication between researchers and businesses working with overseas clients and suppliers. The Web is still not used as a main source of new information largely because most users are not yet adept at searching and the cost of Internet access and download times of web pages discourages both first time and experienced users.

Communicating what?

Websites are now a common feature of large organisations' information services. The website is increasingly seen as an effective way to disseminate and access information in such key areas as health, education, business and government. Most African web content still originates from outside the continent. This trend has to change if web technologies are to make a real and beneficial impact on Africa.

It has been said [6] that African based organisations have to be involved in the production of web content if the technology is to help reduce the information gap found in organisations and different sectors in Africa. In a recent workshop in Zambia, organised by the author as part of an International Institute of Communications Development (IICD) project in partnership with the Internet Studies Research Group (ISRG) at City University, most participants noted that access to information using traditional sources of information was a problem. One participant noted that there is a 'syndrome of institutional selfishness when it comes to sharing information'. People and organisations are not in the habit of sharing information. This largely explains why traditional information resources are poorly maintained as national assets for all to enjoy.

The overall result has been that people place less reliance on local information resources, preferring instead to use external sources. For example, a researcher interviewed prior to the workshop pointed out that he would rather use a market report produced by an external agency based in the UK, the Economic Intelligence Unit (EIU), as such a comparable resource is not available locally. Government departments increasingly refer to overseas sources on information about local conditions.

There is a need for African organisations to take advantage of the Web as it presents an excellent opportunity to catalogue, classify and maintain different types of local information. The author's recent workshops on developing techniques to produce local web content that genuinely reflects the needs of the indigenous people has been welcomed by COMESA (the Common Market for East and Southern Africa) and the United Nations Economic Commission for Africa's Sub-Regional Office for Southern Africa. Both organisations are now participating in a project that involves building web based information resources in the agricultural sector, using a participatory methodology.

The issues about producing web content really concerns projects that are underway and geared towards using websites to disseminate information to people at grassroots level. One of the projects in Zambia, under the auspices of COMESA, involves setting up telecentres where people can access Internet services. Websites targeting local businesses (SMEs) will be built with their direct involvement and a cognisance of their literacy levels. Most of them want to access new information that is jargon free and they are convinced that the Internet is a better way of doing that. They have to travel long distances to town and in some cases they do not get all the information they need because it is widely scattered and is given only to certain people, even if the general public have a right to that information.

Making available to the rural population Internet enabled computers via a telecentre, where one or two individuals have the necessary Internet training, is a model that is being piloted in Zambia and it is already taking root in a number of African countries, such as Ghana, Senegal and Tunisia. Even though rural poverty is a major problem, a large number of people are involved in small scale industries and a large number of them – especially those involved in the production of non-traditional export industries, such as making handicrafts and handmade textiles – want to try to market their wares via the Internet. Only time will tell whether these projects will impact on rural people's income, but the early (and good) news is that people are willing to try the new technologies.

Global developments of ICTs: WAP technologies – Dimension 4

The key to the spread of the Internet is an improvement in the telecommunications infrastructure. Liberalisation of the PTOs, for example, in the Gambia and Ghana has seen the private sector partnering the public sector to improve the service. An even more interesting development saw the establishment of a regional telecommunications company, COMTEL [7], in November last year. Five National Telecommunications Operators (NTOs) and the Common Market for Eastern and Southern Africa (COMESA) established the long-awaited COMESA Telecommunication Company (COMTEL). The formation of this company will see a reduction in payment to European and North American companies of about US$100 per year on transit charges alone on international lines. It is hoped that COMTEL will improve the telecommunication services in the sub-region, extend telephone services in rural areas and provide Internet services in both urban and rural areas.

The growing popularity of mobile telephones in Africa is already solving the voice communication problem. Mobile phones are still a luxury for the urban upper and middle classes. WAP phones are still to appear on the African market and one hopes that mobile phones, designed specifically for use in the Continent, and a hybrid of the Internet that facilitates access first within the national borders and later the region, will bring a taste of the technology to the majority of people who are Internet illiterate. Internet illiteracy is not just limited to the uneducated; there are a lot of university graduates and many office workers in the urban areas that have never used the Internet.

What is, however, noticeable throughout many African countries is the genuine desire to use the Internet whether for business or personal use – and that is highly promising for the future development of the technology. The hunger is there, but what is lacking is the food – the capacity. There is a high level of awareness in countries such as Zimbabwe, where there are a lot of people living overseas who use the Internet to communicate with their relatives back home. These people also play a major role in spreading Internet technologies throughout their countries.

One of the major issues that have dominated discussions on increasing the impact of the Internet is democratising access to the

Internet through the local generation of web content.

The production of quality web content falls within Dimension 3 of the Internet Diffusion Framework used to evaluate the growth of the Internet in Africa. Dimension 1 is concerned with the general utilisation of the Internet in different socio-economic sectors, i.e the application of the features of the Internet. The two key Internet features that are already impacting on communication and information services in many developed countries are email systems and websites. These two Internet services have become a major vehicle for exchanging and accessing information. In Africa, traditional sources of information have proved not to be very reliable, up to date or accurate when compared to information services in many developed countries [8]. The information could be there, but it is not in a readily accessible format as it is scattered throughout numerous localities. The culture of cataloguing, archiving and maintaining library resources generated from within the countries has not been fostered. This has greatly undermined traditional information services and it is only private sector organisations that have collated and maintained their own in-house publications and they do not readily share that information with outside parties.

Government departments are not always keen to freely publicise their activities and there is a general reluctance on their part to keep their citizens fully informed. This is more serious in rural areas where the majority of Africans live. Most of these people are illiterate and sometimes their level of ignorance about basic health issues, for example, has undermined any socio-economic progress in urban towns. Most politicians derive their support from the rural population and it is in their interest to keep them uninformed about the real issues in the country, for that will erode their political power base. From the author's own observation, while on his most recent (2000) field trip in Southern Africa, political institutions are very wary of the power of the Internet to disseminate information freely and independently It is the author's firm belief that proliferation of the Internet in Africa will greatly improve people's understanding of many issues affecting them and it can prove to be a powerful educational tool, equipping people with relevant information that can strengthen their knowledge base and make them truly international citizens. It can be a costly exercise to take the Internet right to grassroots level, but a nation that has a deliberate policy of keeping its citizens fully informed on national issues – social, political and economic – will reap the benefit in the long run.

A case study – web based agricultural information

Field studies, currently being undertaken by the author in Zambia and Zimbabwe, that involve building prototype web information resources in the agricultural sector have revealed yawning information gaps, not just in the agricultural sector but across many key socio-economic sectors, like business, education and health. A recent pilot research project undertaken in Zambia, that involved carrying out informal interviews and filling out of questionnaires on a random sample of 100 farmers, both small scale and commercial, revealed problems in the availability of critical information and access to it [9].

The objectives of the study were to discover:

- what their sources of information are on new agro-technologies;
- how long it takes to get hold of this information;
- the major obstacles in information acquisition;
- how this information affects their productivity on farming on a day to day basis.

The principal findings were:

- farmers whose annual income was less than a $100 had never heard of any new agriculture technologies, and if they did it was a year after release on the market. But those in the income group of more than $1000 had access to the necessary agricultural information;
- 30% of farmers interviewed used the radio as a major source of information;
- 21% used extension workers as a source of agricultural information;

- 7% used the Internet as a source of information;
- the rest confirmed having used more than three sources of information;
- 30% acknowledged having never accessed new information from any of the above sources.

The farmers indicated that they needed information on the following topics:

- farm management, pest and disease control;
- new fertilisers and new seeds (hybrids);
- extension services;
- new irrigation methods.

For those farmers that are export oriented, all of them found it difficult to get trade and market information. The constraints to acquiring new agricultural information included long distances to the information centre and lack of publicity. Most of the farmers interviewed acknowledged receiving information long after it was useful. Of those that accessed information, 80% indicated that there was an increase in the productivity of their farming activities as a result, while the remaining ones did not see any value in agro-technology since they were not interested. Most farmers acknowledged the problem that while they received new agro-technology information it was too complex to understand.

Extension workers assist grassroots farmers by distilling complex agricultural information into a form that is readily understood. It is, therefore, important to supply the latest information to the extension workers, who in turn can empower the grassroots farmers. Carter [10] carried out a major study of printed sources of information used by subsistence farmers in Uganda and Ghana. She acknowledges that most of the farmers are illiterate and rely on each other for new information and usually there are one or two farmers that act as an information focal point for the community. Of course, the problem is that the facilitator can be out of touch and everybody in the community will end up applying old techniques and relying on outdated trade and market information. Most of the farmers interviewed in the author's research project, who came to the Internet awareness workshops, were keen to use the Internet to access new information. They are also keen to participate in identifying the information they need on a day to day, short term and long term basis.

Currently websites on agricultural market information produced from within Africa are thin on the ground. The Southern African Confederation of Agricultural Unions agreed back in 1996 to 'promote the networking of information between organised agriculture in their subregion' [11]. A number of sites have been established in Senegal, Mali, Burkina Faso, Mozambique, South Africa and other SADC countries. The information on these sites is still lacking in terms of content as commercial farmers and outside organisations still do not use the websites as key sources of information, preferring to use printed sources of information that have their own disadvantages, such as being out of date, out of circulation, difficult and expensive to access.

Some of the problems facing the farmers are now being addressed in the IICD project and websites marketing export-oriented farmers' produce will be developed. Trade and market information about prices and overseas producers will also be made available to them via the website. The Zambian project is still in its pilot phase and it will be rolled out to other countries in the COMESA region, and information centres will be established in strategic places throughout Zambia where farmers can tap into this new source of information. Computers will be made available to these information centres and IT workshops will begin in the next six to twelve months (2000/2001).

The way forward

On the basis of this mini study, workshops, conferences, and EInformation Centres will be established in strategic places where the information could be easily accessed as and when it is required. Access is not just about physically connecting to the Internet, although many will enjoy the benefits of using email services both for personal and business purposes. Access is also about improving the quality of web information services. The quality and quantity of locally generated websites is still a major cause for concern. Traditional information services in many African countries have never been developed to an extent which would have facilitated their transfer to the Web, as has been the early pattern in many European and North American countries. Websites, and the information they convey, can fully realise the benefits of Internet technologies on every socio-economic activity in Africa, but someone has to collect and organise this material. The proliferation of Internet technologies is now seen as a yardstick to measure a country's progress towards an information and knowledge based society.

Woods [12], in his article *Content is King*, boldly declares that:

> the creation of knowledge and information is the true driving force behind all the [Internet] hype. The desire to record thoughts, ideas and insights, assimilate those of others will outlive all the current preoccupations with bits, bytes, RAMS and ROMS.

There is a greater need for countries in Sub Sahara Africa (SSA) to generate information resources that reflect their own socio-economic needs (endogenous information), rather than depend on information prepared for them by foreign based organisations (exogenous information). These 'top-down approaches' to producing information ends up stagnating social and economic development [13].

The World Development Report 1998/99, sub-titled *Development and Knowledge*, sees the value of information as critical in the development process. The report also acknowledges the lack of information in many key socio-economic sectors in developing countries and also recommends the pivotal role web technologies can play in collating and disseminating the information. Indigenous people, especially in Africa, should step up their efforts to produce relevant websites that people in their countries can readily access. Visionary leadership is called for to realise the full benefits of new technologies in Africa.

Roy Steiner [14], managing director of Cyberplex Africa, one of Zimbabwe's website development companies, urges Zimbabwe to join the Net revolution. His challenge to Zimbabwe can be equally directed at most of the countries of Sub Sahara Africa. He offers the following suggestions to help get the Internet of the ground·

- *Vision and understanding* of the capacities of online technologies is needed, especially by governments as they can direct investment in the sector and create an enabling environment for the online industry to flourish.
- There is an urgent need to develop the *telecommunications infrastructure*, create a conducive regulatory environment and for financial institutions to take a leading role, in association with the government, to fan the *growth of electronic commerce*.
- 'Silicon valleys', in the form of information technology parks, should feature *as innovative enabling initiatives* to tap into the rich talent in Africa that is being wasted through following educational systems that do not foster graduates to develop original applications of information technology.

African countries should start to shift the focus from land as the mainstay of the economy and harness young talent towards creating IT literate societies, similar to what is happening in India and the Asian Sub-Continent. The teaching of Information Technology subjects should be introduced at all levels in the educational system, as part of a general education programme that will initiate the development of IT creative talent at a very young age. The governments should have IT targets that will show their commitment to build information and knowledge based societies if their countries are to come out of the IT backwaters and enjoy the full benefits of the Internet revolution.

References

1. Press, L. *A framework to characterize the global diffusion of the Internet* 1999. http.//som.csudh.edu/fac/lpress

2. Daly, L. *A conceptual framework for the study of the impacts of the Internet.* 1999. http:// www.bsos.umd.edu/cidcm/papers/jdaly/concept.htm

3. Menou, M. Impact of the Internet: conceptual and methodological issues. In: Nicholas, D. and Rowlands, I., eds. *The Internet: its impact and evaluation. Proceedings of an international*

forum held at Cumberland Lodge, Windsor Park, 16–18 July 1999. London: Aslib, 2000.

4. Jensen, M. *Africa Internet status*. 1999. http://www3.sn.apc.org/africa/afstat.htm

5. UNECA *National information and communication infrastructure – Country Profiles*. Addis Ababa, Ethiopia: African Development Forum, 1999.

6. Chivhanga, B.M. Waking the giant – the Internet revolution in Africa. In: Scammell, A., ed. *i in the Sky*. London: Aslib, 1999, 147–151.

7. COMTEL – a welcome endeavour. *COMESA Journal*, First Quarter, 2000, Lusaka, Zambia, 5.

8. Chivhanga, B.M. The role of the Internet in Africa. In: *Proceedings of the First Annual Global Information Technology Management World Conference*. Memphis, USA: Frogelman College of Business and Economics, The University of Texas, 2000, 171–173.

9. Chivhanga, B.M. Building a web based agricultural information gateway using a participatory approach – a case study from Zambia. *Journal of African Research & Documentation*, SCOLMA (Standing Conference on Library Materials in Africa), London. (In press).

10. Carter, I. *Locally generated printed materials in agriculture: experience from Uganda and Ghana*. Serial Number 31. Department for International Development, 1999.

11. Vallee, J.L. *Market information sources available through the Internet: daily to yearly market and outlook reports, prices and commodities and quotes*. 2000. http://www.aec.msu.edu/agecon/fs2/market/reports_prices.htm

12. Woods, A. The human storage and information retrieval systems. *Business Information Searcher*, 7(1), 1998.

13. Adam, L. Content and the Web for African development. *Journal of Information Science*, 23(1), 1997, 91–97.

14. Steiner, R. Zim must now join information revolution. *Financial Gazette*, August 3–9, 2000, 27.

The automation of controlled vocabulary subject indexing of medical journal articles

David Roberts and Clive Souter
School of Computer Studies, University of Leeds, Woodhouse Lane, Leeds LS2 9JT
daver@scs.leeds.ac.uk

This article discusses the possibility of the automation of sophisticated subject indexing of medical journal articles. Approaches to subject descriptor assignment in information retrieval research are usually either based upon the manual descriptors in the database or generation of search parameters from the text of the article. The principles of the Medline indexing system are described, followed by a summary of a pilot project, based upon the Amed database. The results suggest that a more extended study, based upon Medline, should encompass various components:

- Extraction of 'concept strings' from titles and abstracts of records, based upon linguistic features characteristic of medical literature.
- Use of the Unified Medical Language System (UMLS) for identification of controlled vocabulary descriptors.
- Coordination of descriptors, utilising features of the Medline indexing system. The emphasis should be on system manipulation of data, based upon input, available resources and specifically designed rules.

Introduction

The inclusion of subject indexing terms in the records of bibliographic databases is a common practice. While the sophistication of indexing systems may vary considerably, the objective of this activity is to facilitate effective and comprehensive information retrieval from the databases. The purpose of this article is to consider various aspects relevant to the automation of sophisticated subject indexing. While the utility of such subject indexing may be contentious, the assumption underlying the article is that it does have a useful, even essential, function. A brief overview of some general considerations regarding databases and indexing is presented. Reference is made especially to Medline, one of the best known databases, produced by the National Library of Medicine (NLM), as an exemplar of a database utilising a sophisticated subject indexing system. Next, some of the main lines of information retrieval research are briefly considered with respect to how this research has related to subject indexing. This is followed by a description of the manual subject indexing process, conducted by trained indexers, for the Medline database. The results of a pilot project investigating some aspects of the automation of the indexing process are then presented. These suggest that the possibility of an automated subject indexing system is worth investigating in more detail, though with more sophisticated techniques than were used in the pilot project. Finally, some underlying principles of a more comprehensive approach are suggested. These form the basis of an ongoing project.

Some general considerations concerning indexing for bibliographic databases

Publicly available bibliographic databases are often of considerable size. For example, Medline contains more than ten million records [1]. It is therefore important to ensure that the results of searches on such systems are both relevant and of a manageable quantity.

The utility of subject indexing, based upon descriptors selected from a controlled vocabulary, has been a topic of discussion in the Library and Information Service community over a long period. Nevertheless, there is no doubt that the practice of including subject descriptors in bibliographic records is widespread. The Medline database, originally available only directly by online access to the NLM systems (and those of its agents in various countries), is now widely available through a variety of information service providers in various formats. While some of these organisations originally provided access to versions of the database without controlled subject heading fields, this appears to be no longer the case. It seems reasonable to infer that feedback from users has indicated that use of the subject descriptors can improve the quality of searching. The increasing sophistication of the Medline indexing system can be demonstrated by inspection of the instructions in the Medline Indexing Manual [2] concerning the indexing of, for example, material in molecular biology or haematology. Embase, the main competitor to Medline in the area of general medical databases, currently uses a system of indexing comparable in some respects to Medline. Many databases in other subject fields including, for example, Library and Information Science Abstracts, use some form of subject indexing.

However, the inclusion of subject descriptors in database records has undesirable consequences, including increased production costs and processing time and reduced currency of the database. Where online indexing systems exist, they are usually aids to rather than automation of indexing. Subject indexing generally requires the employment of highly trained teams of indexers. The numbers of indexers employed and the extent of their training and experience may vary with the sophistication of the indexing process, but the cost and time implications are clear. To reduce the cost and time of database production and improve currency seems a desirable objective and automation may contribute to this.

Information retrieval research and subject indexing

The considerable research effort aimed at resolving the difficulties of information retrieval has been comprehensively reviewed by Ellis [3]. This section aims to identify the relationship of the main approaches with respect to subject indexing. Some examples characteristic of these approaches are cited, though not described in detail.

Early experimental studies

The earliest research preceded the wide public availability of online databases, but anticipated the potential of such systems. In the well-known Cranfield studies [4, 5, 6] artificial databases were constructed, which essentially consisted of sets of citations. The objective was to assess the effectiveness of different indexing methods. On the other hand, Lancaster [7] studied the Medline database, attempting to establish searching methods to obtain the best possible retrieval. Both of these sets of studies utilised relevance judgements in assessing system performance. The results suggested (or have been taken to suggest) that for information retrieval by subject, the value of sophisticated indexing systems utilising a controlled indexing vocabulary was doubtful. Such a conclusion was at the time and remains a topic of considerable discussion. One of the principal reasons for the controversy seems to have concerned judgements of relevance which, while intuitively obvious, are subjective and arguably not a basis for objective assessment.

These investigations are considered to have established an experimental approach to information retrieval research, analogous to that of the physical sciences [3]. The information system is conceived essentially as a physical system, the aim being to examine the role of indexing systems in information retrieval. Though the results have raised doubts concerning the usefulness of human indexing activities, these conclusions were and remain controversial.

Statistical and probabilistic approaches

Statistical and probabilistic methods are based upon the implicit or explicit assumption that the subject matter of a document can be expressed in terms of statistical or probabilistic relationships between components of the language used in the document and that these relationships may be utilised in searching. Though all investigations have distinguishing features, a typical example of the statistical approach is the SMART system [8], based upon a discrimination model using term frequency (here 'term' means, roughly, 'word in the text'). It was suggested that terms of medium frequency were those of significance, very frequent terms being insufficiently discriminative and very infrequent terms too rare to be of value for the purpose of information retrieval. A document is characterised by vectors based upon this principle. The search request undergoes a complex analysis to identify similar vectors and retrieval is based upon the similarity between the two sets of vectors.

The underlying idea of the probabilistic approach is that documents should be ranked in order of decreasing probability of relevance to the user. Retrieval is based upon attempts to estimate or calculate this probability. These attempts, as for example by Robertson and Sparck Jones [9], are usually associated in some way with judgements of relevance by the user, with respect to the total number of documents in the database having the property judged relevant.

An idea similar in some respects is that of document clustering, which assumes that closely associated documents tend to be relevant to the same requests. A typical profile of user need for information is produced and the clusters of documents nearest to this profile is retrieved. An early application of clustering to information retrieval was developed by Van Rijsbergen [10].

Essentially, these methods use a different type of representation for user needs and documents or document sets. While interesting, the results of these studies seem not to have established decisively that this approach can be superior (at least in all cases) to other methods of information representation and retrieval.

The cognitive approach

While there are differences of detail between the different methods embodying the cognitive approach to information retrieval, there is a common theme. The cognitive approach is an attempt to model the user need based upon iterative feedback. Thus, in response to user input of some kind, document representations are displayed and the user response results in further refinement of the model to produce a further series of document representations. This continues until the user is satisfied. Thus, these systems were designed to enable the user to navigate the database being searched in an effective way, appropriate to the specific need for information. Relevance is effectively determined by the series of user responses as the process continues.

One of the earlier initiatives of this kind was the THOMAS program described by Oddy [11]. During the dialogue with the user, the program creates an 'image' of the user's perceptions and requirements. It attempts to match this image to images of the documents in the database, defined by a network of associations between documents, authors and subject terms (again, 'terms' means words appearing in the document record) in the database. The document with the image most closely resembling that of the user image is displayed for response by the user, which determines the further refinement of the user image.

A further development of the idea of user need was the concept of 'anomalous state of knowledge', described by Belkin [12, 13, 14]. The chief characteristic of this idea was that the user need is essentially a gap, an anomaly, in knowledge which, as in its very nature this is unknown, cannot be defined. Belkin's system also utilises an association network of concepts to represent the documents in the database and feedback to the user to iteratively define those documents corresponding to the anomaly. A primary characteristic of the cognitive approach to information retrieval is that relevance is determined by the user as the iterative search progresses. However, where complex association networks are created the design effort and system maintenance seems considerable and becomes forbiddingly complex for databases of a realistic size. Also, as the conclusion of the search is determined when the user is satisfied with the output, the adequacy of different forms of subject representation is not necessarily determined.

Expert systems and information retrieval

Expert systems have been used in information retrieval research in an attempt to replace what is usually referred to as the expert intermediary. The expertise referred to in this case is that necessary to construct a search appropriate to the user's needs and the functioning of the system. The components therefore include knowledge of how to interrogate the system, how to identify the appropriate search terms and how to link these by Boolean operators or other methods. While in general users can learn to formulate searches, systems may be of such complexity that methods for obtaining optimal information retrieval are far from self-evident. In practice, much searching even of the most heavily used databases is probably not optimal.

Systems have either focussed directly on the construction of search query statements, which can be used to interrogate the system, or on modelling the interaction between user and database expert. An example of the former is the CANSEARCH system described by Pollitt [15], which was limited to searches of Medline for any aspect of cancer, while MONSTRAT [16] attempts to utilise domain knowledge in the development of retrieval strategies. While in some respects similar to the cognitive approach, these systems tend to be complex in design and construction and may be limited in application, e.g. CANSEARCH is restricted to searches on topics related to cancer. A broader perspective is adopted in GRATEFUL MED, an interface to Medline announced in 1986 by the National Library of Medicine [17] and described by Snow, Corbett and Brahmi [18]. This is now widely known and used. GRATEFUL MED is a front-end to the Medline system and incorporates detailed expert knowledge of how descriptors, assigned to records by specialist indexers using the controlled vocabulary, Medical Subject Headings (MESH), may be utilised in searches.

In general, these systems or interfaces are dependent on the indexing systems of the databases concerned. The implicit or explicit assumption is that the design of the indexing and retrieval facilities is indeed appropriate for the purpose, there being no attempt to evaluate whether or not this is in fact correct. The systems may be supplemented by domain knowledge, i.e. information about the subject area of the database.

Artificial intelligence and information retrieval

The application of artificial intelligence in information retrieval research basically involves processing of source text to identify the roles of words and phrases and the relationships between them. The results of this processing are used to identify appropriate indexing expressions. Natural language processing and semantic and neural networks have been used for this purpose.

Natural language processing

The basic idea is to process the text of documents to generate indexing terms. Methods focusing on the lexical level, attempting to identify grammatical classes or parts of speech of individual words together with machine-readable dictionaries to index documents have, according to Ellis [3], produced 'disappointing results'.

At the syntactic level, there have been attempts to examine the interconnections between words. An example of this is IOTA [19], which uses both co-occurrences of pairs of terms and threshold distances (number of words between two terms). While these notions seem intuitively reasonable, difficulties include the fact that different constructs may be used for the same meaning and no account is taken of word order. Dependency trees have been used in an attempt to resolve the consequent ambiguities, for example by Schwarz [20], further developed by semantic level processing involving contextual knowledge, typified by Berrut [21].

The ideas of co-occurrence and threshold distance may be seen as an attempt to address the issue of coordination (relationships between subject descriptors). Smeaton

[22], commenting on the overall contribution of natural language processing to information retrieval, has suggested that it 'does not really address issues of retrieving information for users based on the language used in queries or texts'. If this opinion were valid, it would seem to imply that while the processing may be ingenious, practical applications for retrieval purposes have not yet been fulfilled.

Semantic networks

This approach is based upon a network of expertise or knowledge, the elements of which are associated with the documents represented in the database. A characteristic example is the ALLOY system, described by Jones et al. [23]. An expert system shell consists of an appropriate conceptual hierarchy and a dictionary of the appropriate subject area. This is created by analysis of subject information by experts in the subject area. Associated with each node in the hierarchy is relationship information for navigating the system and document references. Searching basically consists of browsing the network.

Another example is GRANT [24], which utilised a semantic network of research topics for matching research-funding agencies with researchers. The network nodes were linked by 48 different kinds of relationship, which may be constrained in various ways. The agencies were in a database with a control structure for matching proposals with agencies and an interface for input (the query) and output (agencies matching the query) Compared with an unconstrained search, results obtained by either successively applying constraints (broad first), or successively reducing them, were to obtain more matches but with less precision. ADVISER [25] is a system for monitoring and locating research projects and personnel; initially, subject retrieval relied upon words in source documents, but more recently has involved the creation of a hierarchical classification system.

Neural networks

Neural networks consist of interconnected nodes associated with various kinds of information including document references. However, neural networks differ from semantic ones in that the connections are weighted instead of being 'constrained' and, crucially, the weightings may be modified by user feedback. This consists of relevance judgements on the nodes displayed in response to the initial search enquiry.

An example is Belew's [26] Adaptive Information Retrieval (AIR) system. The system consists of document references, authors and 'keywords' derived from the words in the titles. On input from the user, e.g. a search query, certain nodes are activated and presented to the user. Relevance judgements cause further activation determined by the weightings of the connections between the nodes to be propagated throughout the system. Connection weightings are modified and retained. The process may continue until the user decides the search is complete. In order to establish an effective network considerable input is required from a variety of users.

Some elements of artificial intelligence approaches to information retrieval are similar to some aspects of statistical approaches, the cognitive approach and the use of expert systems. Artificial intelligence approaches do not necessarily exclude data provided as subject descriptors, but the complexity and variety of natural language seems to have resulted in the application of such approaches to information retrieval being far from straightforward.

Summary and conclusions

This brief survey has indicated a considerable research effort into the problems of information retrieval from bibliographic databases and a wide range of approaches. A number of authors have commented on the apparent fact that key problems appear still to await resolution [3, 27, 28]. A common theme seems to be an inadequate theoretical foundation. It is not the purpose in this paper to further contribute to this discussion. Our observations are limited to the relationship of the various approaches to the question of controlled vocabulary indexing by human indexers. It has been a common experience by the providers of large, public bibliographic systems that such indexing is necessary in order to enable a 'reasonable' degree of search effectiveness.

This survey also indicates that information retrieval research has involved various ways of representing the subject matter of documents and various mechanisms for retrieving relevant documents. The relationship between representation type and retrieval mechanisms does not seem to be one-to-one and none of the combinations seems to have achieved general superiority over others. It has been suggested that information retrieval should utilise several different forms of representation [29]. In view of this suggestion – which seems a reasonable one if practicable – and the widespread use of subject descriptors in large public systems, the automation of the assignment of descriptors from controlled vocabularies would seem to be a desirable objective.

Admittedly, this objective, especially in respect to a sophisticated system like Medline, seems of formidable difficulty. A recent preliminary report by Humphrey [30] describes a method of automatic production of controlled vocabulary terms for Medline records. Instead of utilising the full MESH vocabulary this system is based upon the descriptors used for journals in the SERLINE database (consisting of records of the serial collections of NLM). As yet, no retrieval results have been described and whether the use of a general vocabulary for journal articles can be effective remains to be seen. Our own approach is based upon a belief that detailed analysis of textual features of the input can be used to achieve automated indexing using an extensive controlled vocabulary. Before considering this further, a sophisticated indexing system (Medline) is described.

The subject indexing process: a description of key elements and procedures for Medline indexing

Introduction

This account is based on the Medline system for the following reasons: it is widely known and used, it is sophisticated and unambiguously based upon the indexing of concepts, it is (arguably) typical of controlled subject indexing and one of the authors has many years experience in indexing for the database.

The information is presented under a series of headings. In practice, however, the indexing process is interactive in the sense that each of the topics described to some extent informs the others. This 'interaction' tends to increase as indexers become more experienced, the general result being greater speed. Perhaps the most important general ideas regarding indexing are:

- The whole article is not read. To ensure the time spent indexing is not unduly long, indexers must learn to quickly focus on those parts of the article which are crucial. These are usually the title, introduction and results.
- Concepts are indexed only if the article contains significant new information (not including speculative discussion) concerning them. Though articles are not read in their entirety, any section may be inspected for confirmatory purposes or for additional information.
- indexing is based upon the contents of the article not just the abstract (though an abstract of any quality should be an accurate summary of the article).
- Controlled subject indexing is not a word mapping exercise. It is concept indexing, requiring the accurate identification of all significant concepts and their interrelationships and the reflection of this understanding in the accurate selection of coordinated descriptors. The NLM uses the word descriptors (frequently referred to as MESH terms) because this word denotes their function, though they are not, of course, directly derived from the contents of individual texts.
- Indexing is specific. The most specific term available is always used, e.g. HEPATITIS (if correct) not LIVER DISEASES.

Concept identification

The basic instructions for indexers begin as follows. Read and understand the title. Read the introduction. In practice, those parts of the introduction where, for example, work already done relevant to the topic is briefly reviewed, may be only quickly scanned. This material may be of little interest for indexing purposes. However, a sentence beginning 'The purpose of this article is.. ', or something similar, is a key one and should confirm, but may also amplify, understanding of the title.

At this stage, the indexer will probably consult the 'Results' if such a section is included in the article. If not, the portions of the text describing the results must be located. This text may contain important additional information, but only concepts which are 'discussed' are indexed, a 'mere mention' being ignored. How this is interpreted depends considerably on the nature and length of the article or the description of the results. It will be apparent that competence in locating indexable material requires considerable experience. In addition to identifying the concepts the indexer must decide in what context they are discussed, with a view to the correct application of subheadings (described below). For example, is the issue diagnosis, treatment, pathology etc.?

Certain concepts, 'check tags', are an exception to the previous point concerning discussion. Check tags are descriptors very frequently required and include human, animal, common species, age tags (humans), historical period tags (for historical articles). They are always included in the indexing if the relevant information is indicated. The 'check tag' name simply indicates that the terms are listed on the indexing form or online screen and merely require checking rather than typing.

The indexer must decide which of the indexable concepts identified are the most important, which are subordinate and which 'check tag'. The title may be the starting point for this analysis, but some concepts in the title may be subordinate while important concepts may not appear in the title. For example, the title may mention a group of drugs and the article text indicates one or more specific members of this group. The process of identifying the relevant, specific concept is frequently more complex than this example may suggest. A number of additional concepts must also be identified if present. These frequently concern methodology. For example, if a study is stated to have employed a double blind methodology, this is always considered significant, though it may be indicated only briefly in the text.

Finally, the general nature of the article must be determined: editorial, clinical trial (various types), review (various types), letter etc.

Descriptor location

The word 'descriptor' is a synonym for 'MESH term' or 'subject heading', each of which denotes unambiguously a single concept. Frequently, descriptors consist of more than one word (they are pre-coordinated, e.g. LIVER NEOPLASMS). In what follows and in general usage, the word 'descriptor' is understood to include subheadings and check tags, as well as the main contents of Medical Subject Headings. Descriptors are located using a range of indexing tools. The principle ones are:

- Annotated Medical Subject Headings (MESH)· the alphabetic list; all entries indicate the position(s) of the descriptor in the hierarchical structure (the MESH Trees, see next entry) and many entries contain information concerning the specific use of the descriptor. This information is of use to both indexers and searchers.
- MESH Trees: the hierarchical organisation of all the descriptors. While the descriptors are organised into a series of major categories, e.g. category A includes all anatomic descriptors, many appear in more than one position both within and between the major categories. The system may therefore be considered in some respects a network of concepts.
- Permuted MESH: a 'keyword' guide to the descriptors (an alphabetic list of words included in descriptors, each followed by a list of all descriptors containing the word).
- The MEDLARS Indexing Manual: a 300 page reference volume which includes detailed discussion and numerous examples for each of the main categories of descriptors, subheadings, check tags etc. as well as discussions and examples of various specialised areas like microbiology and cancer.
- Supplementary tools: technical notes and

instructions issued from time to time by the NLM.

- Authorities: for Bacteria, Viruses, Fungi, Anatomy, Eponymic Syndromes etc.

Subheadings

Subheadings are about 80 'qualifiers' which are (selectively) attached to descriptors and 'qualify' them (make them more specific). They are very general concepts, e.g. therapy, diagnosis, pathology, and are not used independently (but some of them have equivalent descriptors, used when a general subject or speciality is to be indexed). There are (very complex) restrictions on the application of subheadings to categories and sub-categories of descriptors. The recent trend is for each descriptor to have a unique subset of applicable subheadings. Indexers must assign all applicable subheadings appropriate for the article to each descriptor. Correct descriptor-subheading coordination is a complex aspect of indexing.

Coordination

In addition to the pre-coordination inherent in many descriptors, the main types of coordination are:

a) Descriptor – descriptor: a descriptor implies or requires another descriptor.
b) Descriptor – subheading: applicable subheadings are normally added to all relevant descriptors; some subheadings require a descriptor from another category to be added, e.g. if 'drug therapy' is appropriate for a disease term (category C) then a term for the drug(s) (category D) is also required with the subheading 'therapeutic use'.
c) Descriptor – check tag: for example, a disease term requires 'human' if clinical, 'animal' if experimental and possibly age or species tags.
d) Check tag – check tag: some tags require others, e.g. 'rats' also requires 'animal'.

The indexer must ensure that descriptors are fully coordinated. The effect of this is that the list of descriptors is integrated rather than each being independent, reflecting the relationships between the indexed concepts. While some aspects of coordination are relatively straightforward, some are not.

General

For each concept identified as indexable, the indexer locates the MESH descriptor(s) and appropriate subheadings and check tags using the above tools. As previously mentioned, a distinction is made between concepts considered central to the article and those which are subsidiary. Experienced indexers need the Supplementary tools and Authorities relatively infrequently and can usually immediately identify the appropriate descriptor in Annotated MESH. While indexers may recall frequently used descriptors, it is often still necessary to consult Annotated MESH due to the importance of the information in the annotations.

Appropriate descriptors to indicate the 'type' of article, e.g. editorial, clinical trial, review, letter etc., are added. More than one may be required; journal article is the default, typically for standard clinical and experimental reports. The completed indexing should be reviewed including a comparison of the overall result with the main topic of the article, checking that all necessary coordination is complete and that all necessary check tags have been indicated.

All indexing is subject to 'revision' (by a specially trained, experienced indexer). The extent of revision varies greatly with the expertise and experience of the indexer. A typical example of completed indexing for a relatively short article is given, with comments, in Appendix 1.

Training of indexers

This section is included to further illustrate the complexity of the indexing task. Initial training for new indexers consists of working through about 40 exercises on various aspects. The results for each exercise are reviewed by and discussed with a personal trainer (a reviser). For most people this takes about a fortnight full-time. This is followed by carefully supervised indexing of 'easy' journals, again reviewed and discussed in detail. This may continue for about six months. Individual supervision continues – for about two years, the period necessary for most people to achieve reasonable competence on a wide range of material. Following this, indexing is subject to standard revision procedures. Great emphasis is placed on the 'neutral' attitude to article content by indexers requiring a suspension of any personal criticisms or opinions. An analytic and critical approach to understanding what the article is about (but not necessarily the factual detail) is however clearly necessary.

Final comment

It will be apparent that the primary purpose of the Medline indexing system is the representation of the principal concepts of journal articles. This is done in a way that preserves the main relationships between the concepts, while for each article there are multiple access points corresponding to each concept. This should facilitate a flexible system of combining concepts by an automated information retrieval mechanism. It follows that the primary purposes of such a system will be to provide lists of references about appropriate combinations of topics or concepts. The question of relevance has, of course, been a problematic issue in information retrieval and it is not our purpose here to contribute further to this discussion. However, the following observations seem pertinent. The restriction of relevance to 'aboutness' may be presumed to be a reasonably frequent requirement of databases principally consisting of academic journal articles. The automation of subject indexing based upon 'aboutness' would therefore seem to be a reasonable starting point for investigating the topic

A pilot project in automated subject indexing

A number of considerations seem pertinent to the idea of automating subject indexing The automated selection of indexing descriptors from a controlled vocabulary to a standard comparable with that achieved by human indexers, using sophisticated concept indexing, is an ambitious undertaking. Initially, it may be judicious to limit analysis to titles and abstracts of journal articles. While the use of full text may be desirable, some simplification of the task may be more practical initially and still viable in demonstrating the principle of automated subject indexing. Further, since the overall objective was to identify and index the main and important subsidiary concepts – those central to the overall topic of the article and/or for which new information appeared – a detailed sentence by sentence analysis might not be necessary or appropriate. The system would identify strings of text corresponding to the required concepts. In view of these considerations, a simplified pilot project was considered advisable in order to obtain an indication of whether the methodology may be feasible, before embarking on a lengthy and detailed project. This section contains a brief report of this pilot study.

The database

The database selected for the pilot project was Allied and Alternative Medicine (AMED), produced by the Health Care Information Service (HCIS) of the British Library, for the following reasons:

- The indexing system for this database, while based upon the principles of controlled vocabulary subject indexing is less complex, with fewer descriptors per record, than Medline. In particular, the thesaurus is much more amenable to the development of the system within the constraints of a pilot project, especially concerning time available.
- Records and the AMED Thesaurus were readily available.

Appendix 2 contains examples from AMED records. A general description of the database has been produced by Roberts [31].

The controlled vocabulary

The controlled vocabulary used with AMED consists of a file of about 2,500 records structured as follows·

- MH: the controlled vocabulary term, which constitutes a valid subject descriptor.
- SY: contains synonyms of the MH entry.
- PT: indicates a synonym for which there is a separate MH entry; thesaurus records having this field will not have MH, SY or DE fields since it is necessary to locate the corresponding MH entry.
- DE: a number indicating whether the MH is

to be indexed as a Keyword (KW) or Minor Term (MT). KW corresponds to central concept and MT to subsidiary concept. In general KWs are specific (e.g. disease, drug, organ, organism), while MTs are more general (e.g. they frequently qualify or coordinate with KWs). Some MTs have a similar role to subheadings in Medline but are not specifically attached to KWs.

- CL: a letter indicating the category to which the MH belongs, e.g. A = organ, B = organism, C = disease etc. (The overall categorisation reflects that of MESH.)
- MA: information of relevance to human indexers; this information was not used in the automated system.
- CC· compulsory coordinate; terms also having separate MH entries, which are required if the current MH entry is selected for indexing.

Analysis of titles and abstracts

As article titles do not usually contain verbs, the approach adopted with regard to titles was to eliminate from titles those words without significance for indexing, e.g. definite and indefinite articles etc. Conjunctions and prepositions were regarded as 'concept delimiters'. The analysis results in a list of 'concept strings' which, it was presumed, would be essential to index as they were derived from the title. This presumption is informed by many years of indexing experience and is, in practice, justified with few exceptions. Examples of exceptions include routine technical detail or a specific location of a study. An example of the latter could be 'Leeds General Infirmary' indicating only the incidental fact that the study was carried out there. In general, this kind of information is relatively rare in titles and not relevant for indexing purposes.

The syntactic structure of text in abstracts is more complex than that of titles: abstracts are much longer and frequently contain detail not relevant for indexing, e.g. statistical data. However, indexing based only on the titles is usually insufficient for a complete description of the essential content of the article and the abstract may be a source of important supplementary information even when the title is descriptive and accurate. Two assumptions with regard to the identification of indexable material from abstracts were made. The first assumption was that the appropriate word would appear in the abstract more than once. The second assumption was that an analysis based upon single words rather than 'concepts' would be a sufficient indicator of indexable material, together with the indexing derived from the title, to enhance the indexing accurately. While these assumptions are also informed by indexing experience, they may be not so secure as that mentioned previously with respect to the title. In practice however, with this system, the assumptions seemed to some extent valid. The requirement for analysing abstracts therefore was to eliminate words not required (the same as those not required for the title) and to identify from the remainder those occurring more than a certain number of times. The word frequency, which seemed appropriate, was surprisingly low – three. A number of concepts in the AMED system are analogous to the check tags in MESH, as previously described. These are always indexed, even if only appearing once. The abstracts are searched separately for these.

Finding the controlled vocabulary descriptors

The analysis results in a list of concept strings, each of one or more words located in the title, and a list of words appearing in the abstract at least three times. In what follows the expression 'concept string' refers to components of either kind of list. The general approach to locating descriptors in the Thesaurus consists of a series of operations designated indexing rules. These modify the concept before attempting to locate appropriate descriptors. The indexing rules formulated were:

- Rule 1: concept string unchanged.
- Rule 2: add 's' to the concept string.
- Rule 3: remove the last letter from the concept string.
- Rule 4: interchange the words 'disease' and 'disorder' appearing at the end of a concept string.
- Rule 5: concept string as a component of a descriptor.

- Rule 6: words in a concept string are a descriptor or a component of a descriptor.

The first four rules require an exact match; the last two the best match possible.

Comments on the rules

Rule 1 is self-explanatory. Rules 2 and 3 are intended to account for the majority of cases where the concept string in singular form may be required to map to a plural descriptor and vice versa. While in general the formation of plurals in English is very much more diverse than this, descriptors tend to consist of words that take regular plurals. In practice therefore Rules 2 and 3 are a reasonable approximation of what may be encountered. In an extended system, common irregular plurals could be included if necessary. Rule 4 concerns the use of the words 'disease' and 'disorder' in the medical literature. To be brief, there is in general no consistency – they are used interchangeably. Rules 5 and 6 are used to assemble a selection of possible descriptors for a concept (in the frequent cases where there is more than one possibility) from which a choice may be made later in the processing. The chief arbiter of such choices was the degree of similarity of the concept string to the possible descriptors.

There is no logical guarantee that these rules will produce the optimum result. However, there is also some degree of imprecision in the human indexing process, though the nature of this imprecision may differ from that of the rules used for this system. A principal issue of this investigation was the degree of similarity between the results of the two methods of selecting descriptors.

Implementation

The system was developed in a Windows NT environment using Visual Basic The records and the thesaurus are accessible as MS Word files and output is either displayed on the screen or output to MS Word for detailed comparison with human indexing of the same records. This is a convenient arrangement as MS Word can be run directly from the application. Visual Basic also incorporates an extended series of commands appropriate for word processing manipulations, e.g. finding, copying and pasting sections of text. Other information, including lists of prepositions and conjunctions, was built directly into the application in array structures. Processing time was considerably extended when indexing Rules 5 and 6 were called by the system, but seemed to be two to three minutes per record on a typical Pentium desktop personal computer.

The indexing process: illustration and results

Before presenting a summary of the overall results, a specific example is described:

Title: *Multiple sclerosis presenting as spinal cord disease*

Concept strings: *Multiple sclerosis presenting Spinal cord disease*

This simple title is split at the word 'as' The word 'presenting' is not included in the words eliminated during the analysis and is therefore included in the concept string. In this example (and many similar occurrences in the records processed) such circumstances do not necessarily prevent location of the appropriate descriptor.

Descriptors located based on the title:
Keywords: MULTIPLE SCLEROSIS,
SPINAL CORD DIS
Minor Terms: (none)

The word 'presenting' has, in effect, been ignored though all the indexing rules would be called to achieve this result. The correct 'DIS' descriptor has also been found; there is a special procedure which abbreviates all 'disease(s)' at the end of a concept string as all such descriptors in the AMED system are abbreviated in this way (but not if 'disease' occurs elsewhere in a descriptor). The abstract for this record follows.

> Multiple sclerosis, a disorder of central nervous system demyelination, is a leading cause of disability in young people. Lesions of the spinal cord are usually less than two vertebral body segments long, peripherally located, and found in the cervical region. An adult woman had a 2-month history of back pain, urinary incontinence, and bilateral lower extremity weakness. Magnetic resonance imaging (MRI) of

> the spine showed an intramedullary spinal cord tumor from T4 to T8 with an intramedullary cyst from T1 to T4. After thoracic decompressive surgery, findings from biopsy of the cord lesion were consistent with multiple sclerosis. Postoperatively, the patient required an intensive rehabilitation program. This is the first reported case of histopathologically confirmed spinal cord demyelination presenting as an intramedullary thoracic cord mass. Physiatrists should be alerted that demyelination disease can mimic a spinal cord tumor, even on MRI, and must be considered in the differential diagnosis of a symptomatic spinal cord mass.

Words listed from the abstract by the system were:

demyelination
intramedullary
spinal
cord

Words not of significance for indexing (prepositions, conjunctions etc.) are excluded from the list. Also, words corresponding to check tags are also excluded though the appropriate descriptors are added as Minor Terms.

The complete list of descriptors for this record:

Keywords: MULTIPLE SCLEROSIS, SPINAL CORD DIS, DEMYELINATING DIS
Minor Terms: ADULT, CASE REPORT, FEMALE, PATHOLOGY

The indexer assigned the following descriptors to the record:

Keywords: MULTIPLE SCLEROSIS
Minor Terms: SPINAL CORD DIS, NERVOUS SYSTEM NEOPLASMS, CASE REPORT, DIAGNOSIS DIFFERENTIAL

The automated system has added the Keyword, DEMYELINATING DIS, and four Minor Terms. Based upon the abstract, the selection of DEMYELINATING DIS seems reasonable and may be regarded as an improvement by the automated system. The word 'demyelination' has also resulted in the addition of the compulsory coordinate, PATHOLOGY. The first three of the additional Minor Terms are all required if they occur once (the 'Class' field of the appropriate thesaurus entry indicates this requirement). The inclusion of 'spinal' and 'cord' in the list of words have no effect as they both occur in a term already selected, SPINAL CORD DIS. There are no descriptors in the AMED Thesaurus containing the word 'intramedullary' so this word produced no result.

The human indexer included only MULTIPLE SCLEROSIS as a Keyword with SPINAL CORD DIS as a Minor Term. Also as a Minor Term was DIAGNOSIS DIFFERENTIAL though ADULT and FEMALE were omitted.

As the 'spinal cord disease' concept appeared in the title, it would seem more appropriate to index it as a Keyword, not as a Minor Term. The automated result in this respect would therefore seem preferable. The use of the term DIAGNOSIS DIFFERENTIAL is specially defined within the rules of the indexing system. It is for the situation where a health care practitioner is trying to distinguish between two or more conditions, i.e. differentiating the diagnosis. The expression 'presenting as' in the title is a characteristic indicator of this situation. The human indexing is in this case correct within the scope of the indexing rules. This is, of course, an example of just the kind of circumstances that are especially difficult to design a computer system to process correctly.

The omission of the terms, ADULT and FEMALE, is a definite error by the human indexer and an example of the kind of circumstances in which a computer system can be more efficient than a human indexer.

Finally, the indexer assigned the term NERVOUS SYSTEM NEOPLASMS. This is implied by the use of the words 'tumor' and 'mass' in the abstract and may be clarified and confirmed by the full text. This is an example where arbitrary processing rules have resulted in an omission of a coordinate descriptor and illustrates the difficulty of designing such rules to simulate comprehensively human processing. A facility to do this would, in principle, be possible utilising the hierarchical organisation of the descriptors, but this was not developed in this preliminary project.

Overall, for this record, there are positive and negative features with respect to the performance of the system as judged by comparing the indexing with that of a human indexer. This is emphasised and reflected in the results of processing a series of records. These results are now summarised.

Results of processing

Selection of records

It was necessary to impose certain constraints on the selection of records for processing·

- A descriptive and precise title is necessary since the analysis of the title is the main way of identifying the principle concepts. The underlying (and usually reasonable) assumption is that titles of articles in professional journals will usually meet this requirement. Articles with very general or misleading titles were therefore excluded from those that were processed. This may be viewed as a bias in the sample or a constraint on the capability of the system.
- All articles selected must have an abstract since processing of a title alone is unlikely to be adequate. Sometimes very long abstracts contain numerical and statistical data and other details not relevant to the indexing process. To limit the processing time such material was removed from the abstracts concerned. This does not affect the operation of the system but merely reduces the quantity of material to be processed.
- Titles and abstracts must be in the English language
- Articles on specialities within the medical domain may present a particular problem for automated indexing. While appropriate concepts may be identified, they are infrequently of a general nature, e.g. rehabilitation, physiotherapy, and may therefore be allocated as Minor Terms rather than Keywords. The approach was not to exclude such articles but to ensure they are not over represented as a proportion of the articles processed.

Subject to these constraints an attempt was made to include in the articles selected for processing a range of subject areas, e.g. rehabilitation, physiotherapy, palliative care, physiology, biomechanics. We estimate that at least 80 per cent of articles in AMED (excluding letters and editorials) are standard articles with titles of good quality and that any bias in the selection of records would not significantly affect our overall conclusions regarding the possibility of developing automated indexing.

Results

The numbers of descriptors for human and automated indexing for 100 processed records were as follows:

	Keywords	**Minor Terms**	**Totals**
Human indexing	290	210	500
Automated indexing	370	180	550
Both methods	199	38	237
Automated – added correctly	49	63	112
Automated – added incorrectly	122	79	201
Automated – omitted	91	172	263

The records were processed in two batches of approximately equal size. Though there were some differences between the batches, the figures were broadly comparable so the table shows data for the two batches combined. The comparison was initially based on an assumption that descriptors included by human indexers were correct. Indexing errors in AMED are usually omissions rather than incorrect inclusion of descriptors. Erroneous addition of descriptors by the automated system was usually evident from the inspection of the titles and abstracts. The main difficulty, therefore, in the comparison was in relation to descriptors included by the automated system and omitted, apparently incorrectly, by human indexers. A cautious approach was adopted in looking for evidence in titles and abstracts (it was assumed a descriptor was added erroneously unless evidence suggested the contrary). While there is, of course, subjectivity in judging that a descriptor should have been included, as mentioned previously, these judgements were informed by indexing expertise In many cases, as in the above example, these judgements seemed reasonably secure. A notable trend is that the automated system tends to assign more Keywords and fewer Minor Terms.

The total number of errors was 464 (4.6 per record). Against this, the system was judged to have selected 112 (1.1 per record) descriptors omitted by indexers, but which were appropriate. Subtracting the latter from the former gives a net difference figure of 3.5 per record; which may be considered as an indicator of net error rate. Ideally, this figure would be zero or negative, indicating that automated indexing is as good as or better than human indexing.

The errors in the automated indexing were examined to try to determine the cause. For omitted terms (both Keywords and Minor Terms) a classification was possible as follows:

Coordination: 147	Semantic, e.g. synonyms: 25
Not apparent: 59	Word form e.g. adjectival form: 32

Coordination generally involves adding descriptors to create intellectual links between selected descriptors. Compulsory coordinates are indicated in the CC field of the thesaurus. Coordination requirements, however, often cannot be specified in this way, or are discretionary. The 'Not apparent' category includes cases where it seems that the system should have produced the term omitted, but for some reason did not. Such cases merit further investigation in an extended project.

There were no detectable trends or patterns in the descriptors incorrectly included by the automated indexing. In general, these errors seemed to be related to the analysis process rather than descriptor selection.

Discussion

The results from the processed records, summarised above, can only be regarded as a general indication of the quality of indexing possible in an automated system. Reasons for this include:

- Only a relatively small number of records were processed (there are currently about 85,000 in AMED), for various practical reasons. It seems likely, however, that the number processed was adequate to establish that the system is working broadly as intended.
- There may be some bias in the selection of records, though an attempt was made to achieve a broad and 'typical' selection, subject to the capacity of the system.
- Various aspects are incompletely developed. It was always recognised that the analysis of the input (titles and abstracts) would be relatively unsophisticated and that a complete system would require more attention to coordination of the descriptors. A further desirable feature would be some kind of internal feedback mechanism to monitor and modify the contents of the controlled vocabulary, according to the contents of the literature being processed.
- The broad picture indicated is that the automated system produced on average 5.5 descriptors per record, of which 3.5 were correct and 2.6 omitted. Based upon descriptors included apparently correctly by the system but omitted by indexers, there should be on average 6.1 descriptors per record. While the automated system has considerable deficiencies, in some respects it represents an improvement.

The results may be considered to suggest that in relation to what is required the automated system was rather more than 50 per cent correct. However, it is clear that a viable automated indexing system would have to achieve considerable improvements both in terms of identifying correct descriptors and omitting incorrect ones.

The purpose of bibliographic databases, the ability to find information relevant to a need, to perform effective information retrieval, has not been directly addressed. An underlying assumption has been that a system based upon controlled indexing and concept mapping can result in effective information retrieval. The immediate objective was to establish if a more substantial research effort in automated subject indexing has some prospects of success. The evaluation of such a system would involve not only an inspection and comparison of the indexing produced, but also an attempt to determine its information retrieval performance. The next section outlines some principles, which may inform a more sophisticated approach to automated concept indexing.

A sophisticated automated indexing system: principles

In our view, the results of the pilot project are sufficiently encouraging to suggest that the possibility of an automated indexing system deserves serious investigation. This should be informed by the following considerations.

1. Language is to be considered as a tool used for specific purposes, and having characteristics corresponding to that purpose. In this case, the purpose is the presentation of the results of clinical and experimental research to an audience of researchers and others interested, i.e. people who read the journals. This broadly follows Wittgenstein's later views, as described by Blair [27] and Brier [28].
2. An analysis of the features and regularities of the language characteristic of documents to be represented will be indispensable. Biber [32] has described an extended investigation of language features that can be used to classify the genre of natural language sub-corpora. While, for the purpose of indexing medical journal articles, the necessary features and regularities required are likely to be different from those determined by Biber [32], the investigation would be in the spirit of his approach.
3. The principle task of the system is the representation of documents [27]. This is to be done in a way appropriate for the purpose, specifically regarding subject content. A primary characteristic of the representations will be the preservation of essential elements of the context of word uses in the literature. Aspects of the system supporting the identification and preservation of context in the representations include: restriction to a domain (both subject and type of text), analysis of texts for features and regularities, coordinated indexing vocabulary and integration of the indexing descriptors (coordination).
4. Following the critique of Frohmann [33], a mentalistic conception of indexing is rejected. Indexing 'rules', for both human and automated indexing, are created for the specific purpose. The representations in a computer system in no way imply an 'understanding' of the material. This is a matter for human beings, with respect to creation of original documents and design and use of automated systems.
5. Development embodies the explicit recognition that algorithmic computer processing does not necessarily simulate human cognitive processes. The system produces representations of documents which humans can use for specific, defined purposes.
6. The system will be based upon a suitable, controlled subject vocabulary as its primary 'lexicon'. This should be based upon the practice of indexing rather than be some theoretical construct Important advantages of the vocabulary include the use of coordinated descriptors and the elimination of semantic ambiguity with respect to the domain. In medicine, the Unified Medical Language System [34], an extensive consolidation of medical vocabulary systems, produced by the National Library of Medicine, seems an appropriate choice.

Conclusion

While the use of controlled subject indexing in bibliographic records has been a topic of considerable discussion, controlled vocabulary descriptors do seem to be valued in searching. Attempts to devise alternatives have met with limited success, at least with respect to major public access databases. The discussion may be influenced by an awareness of the growing difficulty of maintaining such features. Aspects of these difficulties include the ever increasing number of documents published, the great and continuing improvements in the physical (but not intellectual) accessibility of documents through computer systems and the relative scarcity and cost of indexing expertise. A further factor may be the perceived difficulty of devising an automated method of producing subject indexing of a quality comparable to that of human indexing. The sophistication and complexity of indexing for Medline has been illustrated in this article. While the automation of subject indexing is undoubtedly a task of formidable difficulty, this article has suggested that a serious investigation of this possibility may prove fruitful, at least with respect to original experimental and clinical reports in medicine. The pilot project described a relatively simple approach to the automation of the indexing for a small and relatively specialised database. In the light of the known deficiencies and constraints, the results obtained seemed modestly encouraging. The final section of the paper outlined some basic principles of further development with the objective of much improved results. Particular characteristics of the approach are identification of features and regularities of linguistic usage of texts within the domain and the use of an extended 'lexicon' of medical vocabulary (UMLS). While drawing from a variety of sources, UMLS has been integrated into MESH, which has resulted from many years of subject indexing and may, therefore, be a tool appropriate for the task. This extended project is ongoing.

References

1. Personal communication, 1999.

2. National Library of Medicine *MEDLARS Indexing Manual*. Available at: http://www.nlm.nih.gov/mesh/99MBrowser.html

3. Ellis, D. *Progress and problems in information retrieval*. London: Library Association Publishing, 1996.

4. Cleverdon, C.W. *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*. Cranfield: College of Aeronautics, 1962.

5. Cleverdon, C.W., Mills, J. and Keen, E.M. *Factors determining the performance of indexing systems, (Volume 1: Design)* Cranfield: College of Aeronautics, 1966.

6. Cleverdon, C W., Mills, J. and Keen, E.M. *Factors determining the performance of indexing systems, (Volume 2· Test results)*. Cranfield: College of Aeronautics, 1966.

7. Lancaster, F.W. *Evaluation of the MEDLARS demand search service*. Bethesda: National Library of Medicine, 1968.

8. Salton, G. *The SMART retrieval system – experiments in automatic document processing*. Englewood Cliffs: Prentice Hall, 1971.

9. Robertson, S.E. and Sparck Jones, K. Relevance weighting of search terms. *Journal of the American Society of Information Science*, 27(2), 1976, 129–146.

10 Van Rijsbergen, C.J. *Information Retrieval*. 2nd edition. London: Butterworth, 1979.

11. Oddy, R.N. Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1), 1977, 1–14.

12. Belkin, N.J. Anomalous states of knowledge as the basis for information retrieval. *Canadian Journal of Information Science*, 5, 1980, 133–143.

13. Belkin, N.J., Oddy, R.N. and Brooks, H.M. ASK for information retrieval I· background and theory. *Journal of Documentation*, 38(1), 1982, 61–71.

14. Belkin, N.J., Oddy, R.N. and Brooks, H.M. ASK for information retrieval II: results of a design study. *Journal of Documentation*, 38(2), 1982, 144–164.

15 Pollit, S. CANSEARCH: an expert systems approach to document retrieval. *Information Processing and Management*, 23(1), 1987, 119–138.

16. Belkin, N.J., Seeger, T. and Wersig, G. Distributed expert problem treatment as a model of information system analysis and design. *Journal of Information Science*, 5(5), 1983, 153–167.

17. National Library of Medicine. Grateful Med. *NLM Technical Bulletin*, 205, 1986, 11–12.

18 Snow, B., Corbett, A.L. and Brahmi, F.A. GRATEFUL MED: NLM's front end software. *Database*, 9(6), 1986, 94–99.

19. Chiaramella, Y. and Defued, B. A prototype of an intelligent system for information retrieval: IOTA. *Information Processing and Management*, 23(2), 1987, 285–303.

20. Schwarz, C. Content based text handling. *Information Processing and Management*, 26(2), 1990, 219–226

21. Berrut, C. Indexing medical reports: the RIME approach. *Information Processing and Management*, 25(1), 1990, 93–110.

22. Smeaton, A.F. Progress in the application of natural language processing to information

retrieval tasks *The Computer Journal*, 35(3), 1992, 268–278.

23. Jones, L.P., deBessonet, C and Kundu, S. ALLOY: an amalgamation of expert, linguistic and statistical indexing methods. In: Chiaramella, Y., ed. *Proceedings of the 11th. International SIGIR Conference on Research and Development in Information Retrieval*. NewYork: ACM, 1988, 191–199.

24 Cohen, P.R. and Kjeldsen, R Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management*, 23(2), 1987, 255–268.

25. Dew, P., Leigh, C and Whyte, W *ADVISER II: theory and practice of finding and representing RTD results*. Presented at CRIS 2000, Knowledge at work – research information for society, Helsinki, 25–27 May 2000.

26. Belew R.K. Adaptive information retrieval using a connectionist representation to retrieve and learn about documents. In· Belkin N.J. and Van Rijsbergen C.J., eds. *Proceedings of the 12th International SIGIR Conference on Research and Development in Information Retrieval* New York: ACM, 1989, 11-20.

27 Blair, D.C. *Language and representation in information retrieval* Amsterdam: Elsevier Science Publishers. 1990.

28. Brier, S. Cybersemiotics: a new interdisciplinary development applied to the problems of knowledge organisation and document retrieval in information science. *Journal of Documentation*, 52(3), 1996, 296–344.

29. Croft, W.B., and Turtle, H.R. Text retrieval and inference. In: Jacobs, P.S., ed. *Text-based intelligent systems: current research and practice in information extraction and retrieval*. New Jersey Lawrence Erlbaum Associates, 1992, 127–155.

30. Humphrey, S.M. Automatic indexing of documents from journal descriptors: a preliminary investigation. *Journal of the American Society of Information Science*, 50(8), 1999, 661–674.

31. Roberts, D.J. AMED: a bibliographic database for complementary medicine and allied health. *Complementary Therapies in Medicine*, 3(4), 1995, 255–258.

32. Biber, D. *Variation across speech and writing*. Cambridge· Cambridge University Press, 1988.

33. Frohmann, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, 46(2), 1990, 81–101

34. National Library of Medicine. Fact Sheet. *Unified Medical Language System*. 1999. (also available at http://www.nlm.nih.gov/factsheets/umls.html)

Appendix 1. An example of completed indexing

TACHYCARDIA, VENTRICULAR / * drug therapy
ANTI-ARRYTHMIA AGENTS / * therapeutic use
PREGNANCY COMPLICATIONS, CARDIOVASCULAR / * drug therapy
TACHYCARDIA, VENTRICULAR / physiopathology
PREGNANCY COMPLICATIONS, CARDIOVASCULAR / physiopathology
HEART VENTRICLE / physiopathology
ELECTROCARDIOGRAPHY
PREGNANCY (Check tag)
HUMAN (Check tag)
FEMALE (Check tag)
ADULT (Check tag)
DOUBLE BLIND METHOD
RANDOMIZED CONTROLLED TRIAL [Publication type]

Comments:

a) This is a clinical trial of ventricular tachycardia in pregnant patients treated by one or more anti-arrythmia agents (ventricular tachycardia is a type of arrythmia). The author has stated that a double blind methodology was used.

b) * indicates a primary or central concept.

c) Note the 'matching' pair of subheadings (drug therapy and therapeutic use).

d) PREGNANCY COMPLICATIONS, CARDIOVASCULAR is a required coordinate, bringing together the concepts pregnancy and the disease condition. Such coordinates take all appropriate subheadings and are required to facilitate effective and comprehensive searching.

e) The authors presented information on aspects of ventricular function in these patients (function in a disease is physiopathology) at least partly based on electrocardiographic investigations. Note that in addition to the organ (HEART VENTRICLE) the subheading is added to the relevant disease descriptors.

f) ELECTROCARDIOGRAPHY is added to indicate a method used. The subheading 'methods' is only added if the specific method (of ECG in this case) is a particular issue (for example, if there is something different about how the ECG was done). The criteria for adding technique descriptors to indicate a method used routinely are somewhat flexible and seem to depend more on indexer experience than precise rules or guidelines.
g) 'Check tags' are always required if the relevant information is given. A mention is sufficient. (Check tag indexing may be viewed as more like classification than descriptive indexing.)
h) The study is described as a randomised trial using the double blind method. The indexer does not attempt to evaluate if this is correct – it is assumed that what is reported is true. Whether the methodology is of sufficient rigour to justify the description as a double blind randomised trial is the responsibility of the authors, peer reviewers and journal editors.
i) This quantity of indexing is likely to result from a fairly short account, e.g. a typical contribution of four or five pages in the British Medical Journal. Articles may require more terms or fewer. For example, the indexing of Brief Reports, Case Reports, letters etc. may be relatively economical.

Appendix 2. Some AMED Records

AN: 0005933
AU: Fullerton BD, Browngoehl LA
TI: Total knee arthroplasty in a patient with bilateral Charcot knees
SO: Arch Phys Med Rehabil 1997 Jul; 78(7):780–2
ET: Charcot Marie disease
KW: ARTHROPLASTY, KNEE JOINT, ARTHROPATHY NEUROGENIC
MT: CASE REPORT, SURGERY OPERATIVE
AB: Neuropathic arthropathy (Charcot joint) is a progressive and degenerative process resulting from underlying neurovascular and neurotraumatic deficits. Diabetes mellitus is now the most common cause of Charcot joint. The authors report a case of a 61 year old diabetic woman with bilateral Charcot knees who successfully completed a rehabilitation program and achieved independence after left knee arthrodesis and right total knee arthroplasty.

AN: 9159144
AU: Baguley IJ, Felmingham KL, Lahz S, Gordan E, Lazzaro I, Schotte DE
TI: Alcohol abuse and traumatic brain injury: effect on event-related potentials.
SO: Arch Phys Med Rehabil 1997 Nov, 78(11):1248–53
KW: ALCOHOL DRINKING, BRAIN INJURIES, EVOKED POTENTIALS
MT: PHYSIOPATHOLOGY, NEUROPSYCHOLOGICAL TESTS
AB: OBJECTIVE: To examine the individual and combined impact that traumatic brain injury (TBI) and heavy social use of alcohol have on electrophysiologic correlates of working memory and evaluation of task-relevant information. DESIGN: Case-control study. SETTING: University hospital brain injury rehabilitation unit. PARTICIPANTS: Forty male volunteers divided into four groups on the basis of their history of TBI and alcohol intake. Subjects with TBI had experienced a severe closed head injury at least one year before testing MAIN OUTCOME MEASURE: Event-related potentials (ERPs) and neuropsychometric tests. RESULTS: Groups showed no significant differences in average age or neuropsychological tests. TBI groups did not differ in time postinjury or on severity measures. Alcohol use measures were significantly greater in the two alcohol groups. N200 latency and P300 amplitude were impaired in heavy social drinkers and in nondrinking subjects with TBI relative to controls, but were significantly impaired in subjects with TBI who were also heavy social drinkers. CONCLUSION: The results indicate that although alcohol use and TBI independently produce mild alterations in some aspects of late ERP components, the ERP changes are significantly greater when alcohol use and TBI are combined. This study provides evidence that heavy social drinking after TBI has a measurable impact on electrophysiologic correlates of cognition.

Narrative knowledge: eliciting organisational knowledge from storytelling

Stuart Hannabuss
School of Information & Media, Robert Gordon University, Aberdeen
s.hannabuss@rgu.ac.uk.

Knowledge is critical for organisational effectiveness and competitive advantage. Knowing that you know, knowing what you know, and knowing that you do not know, are critical aspects of knowledge management. Increasing emphasis is placed on the need to identify and use tacit knowledge, as well as explicit knowledge. This discussion examines the unique role of narrative (in the form of storytelling) in eliciting tacit knowledge (including tacit meta-knowledge) in the sensemaking of organisations.

Knowing, competence and meta-knowledge

Knowing consists of many states. These include knowing that you know, knowing that you do not know, not knowing that you know, and not knowing that you do not know. Each state has its own role to play in decision making in organisations and in daily professional life. It might be said that knowing that you know is *a state of conscious competence*, where a manager knows he/she can do things and can do them. Knowing that you do not know is a good starting point for learning and training, and asking for expert help. This is *a state of conscious incompetence.*

Not knowing that you do not know presents more challenges because ways may need to be found to change not knowing into knowing about your ignorance. States of not knowing may arise innocently or through an indifference to learn and update knowledge and skills, and there may be elements of shame or embarrassment when not knowing turns (or is turned) to knowledge. Not knowing that you do not know leads to poor decisions and self-deception. Many organisations think they are doing well and then fail. This is *a state of unconscious incompetence.*

Then there is not knowing that you know. This is *a state of unconscious competence.* The knowledge, or capacity for knowledge, or latent ability to recognise and use knowledge, is actually there. It is just that the manager or organisation does not know it is there. In fact, an organisation may have access to extensive forms of encoded information and knowledge, for example, in its legacy systems. Based on these, decisions, analyses and forecasts can be made.

In knowledge management the state of unconscious competence is particularly interesting because it is a continuum. At one end it consists of not knowing that you know, in the sense that there is some implied deficit or shortfall. Not knowing deprives you of opportunities to make informed choices, perhaps about the future. Not knowing means that you fear rather than embrace tolerable risks, that you place conservative reliance on information management and the systems that support it in the belief that what is there is what is needed, and that it is enough.

In the middle of the spectrum of unconscious competence there is the unconscious competence of the concert pianist and chess master and natural leader: one whose skills are so implicit, so much a natural and unconscious part of performance that the constituent skills contributing to playing a piano concerto or mating an opponent has a large unconscious element. Skills have become internalised into a cognitive and attitudinal script in the person concerned. Even they may find it difficult to explain how they can do things. In fact to think of fingering, even to think of memorising, while playing the concerto might be to interrupt the unconsciously competent mastery of the art. For organisations, this stage indicates an ability to function well, make good decisions, and trust in its systems and people.

At the other end of the spectrum of unconscious competence we find that the unconscious part is evolving into a meta-conscious part. Following the logic of the reflective practitioner and meta-cognitive learning in training, meta-learning is where we acknowledge and use our learned knowledge of learning to help learning, arguably making learning richer and more personalised. Knowledge and competence are particularly self-aware processes, so it is no surprise when we apply concepts of meta-learning to them. Taking steps to reflect on how conscious we are about our competence makes sense.

It leads to a clearer understanding of where we are competent and why we think or believe we are competent. It also leads, or should lead, to a clearer mind-state in which we can start to identify what we most need to develop and put in place in order to ensure that managers in the organisation do not just have the information they need, or the knowledge they need, let alone the knowledge they do not have or do not know they do not have – but also the knowledge of why and how they can be unconsciously competent. This in turn picks up the importance of tacit knowledge which commentators from Polanyi to Nonaka have described. There is conscious tacit knowledge – we know or think it is there, in the organisation as a whole, or project team, or an individual manager's head.

Mining tacit knowledge is a mantra of modern knowledge management [1, 2]. But below (if we accept the metaphor of levels, like soil or the oceans) conscious tacit knowledge (itself below explicit knowledge) there is unconscious tacit knowledge. It may well 'cover' or 'relate to' the same or similar areas of knowledge and understanding, memory and logical inference, intuitive connectedness and foresight which conscious tacit knowledge covers and relates to (and by that token, explicit knowledge openly deals with).

Unconscious knowledge needs to be made conscious before it can be used. Tacit knowledge must become explicit before it can be used. Therefore, there is a real challenge in making use of unconscious tacit knowledge. Meta-knowledge occurs at, and enables, the transitions to be made between explicit and tacit, and between conscious and unconscious tacit knowledge, growing ever more patent or overt at each stage.

Knowledge Process

KNOWLEDGE TYPE	PATENCY	META-KNOWLEDGE
explicit	conscious	
		transition
tacit	conscious	
		transition
tacit	unconscious	

The differences between conscious tacit knowledge and unconscious tacit knowledge are interesting. We are conscious that the first is there. It hovers about explicit knowledge, providing a personalised, impressionistic, metaphorical and symbolic hinterland for the many transactions and choices made in the organisation, corporately, interactively, and individually. The assumption is that, when and where it is made explicit, it is going to be of some use to the process of work, however subjective or idiosyncratic it is.

We are far from conscious that the second is there because it is unconscious. The fact that it is unconscious may arise because we have no knowledge of its existence, or because we have become so familiar with it that we no longer see its existence. We realise its existence quintessentially in what is called an 'aha!' experience, when we encounter something or we are told something, and we recognise it for what it is with an intuitive naturalness and speed which often surprises us. This act of recognition might derive from having seen it, or major elements or patterns of it, before. On the other hand, recognition can come when we see that two and two make five and have done so all along.

It is at that point, too, when we know that we know, know that we have not known fully until then, recognise the interconnectedness of past and present, ourselves and the other, that we become strongly aware of the meta-cognitive aspect of tacit knowledge use. When unconscious tacit knowledge becomes conscious tacit knowledge, that is the point at which awareness hits us: that is the tacit meta-knowledge point For individual thinkers, this is a critical moment of awareness – part of the process and mystery of creativity and innovation.

For organisations it is the point where people realise they have the capacity for great ideas, making the link. It is a configurational process, where cognitive gestalts reveal hidden truths about connections [3, 4] Finding ways of identifying and eliciting knowledge, and knowledge about knowledge, is likely to be one of the greatest challenges for, and hallmarks of, an effective organisation.

Dealing with tacit knowledge in itself presents organisations and research teams with ethical challenges too, above those to do with informed consent, surveillance in the workplace, intellectual property rights, and personal privacy and identity. What is tacit may, for political or proprietary reasons, be regarded as needing to be kept so. The rhetoric of 'corporate memory' glibly buries the humanistic diversity of people at work under a credo of conformism and employee transactional utility.

So ways need to be found of eliciting knowledge which also disinterestedly and authentically allows workplace stakeholders to share knowledge as equal players. Research into organisational cultures reveals time and time again preoccupation with issues of power, control, and equality. Taylorian models of management historically typify this imbalance of power, turning much management research into a form of institutionalised panopticon. This is what Foucault objected to.

In modern times, with a wider range of transactional and flexible contractual arrangements between employers and employed, it seems an appropriate time to emphasise ways of identifying and eliciting knowledge which attracts belief and trust, and which grows organically from a natural humanistic soil. This discussion examines ways in which narrative, and storytelling, enables managers, researchers, and organisations to achieve this.

Sensemaking and identity

Explanations which assume that organisational behaviour is wholly rationalistic (e.g. that decisions are made in states of full objectively gathered knowledge, that managers are always reasonable) usually tend to omit the ethnographic complexity of the workplace and interactions there [5, 6]. It is for this reason that in research on organisational cultures in recent decades there has been a strong emphasis on anthropological and ethnographic approaches.

In the first case, there is a recognition that organisations, like other social groups, reveal things about their customs and rituals, symbols and transactions when observed and studied in the manner of an anthropologist. Participant-observation and insider-research using case studies typify this approach. In the second case, there is a further recognition that people make sense of situations, events and states in important and investigable ways: the discourse they use to describe these situations, events and states is researchably part and parcel of sensemaking, by the researcher as well as by all stakeholders involved.

Ethnographic research emphasises the ways in which the discourses, rhetorics, texts, metaphors, images and symbols play a critical mediating role in sensemaking. Even though criteria might be established in advance, interview schedules drawn up, and specific outcomes identified, nevertheless it is regarded as important to examine not just what happens but what the actors thought happened, what led up it, what caused it, how participants acted, whether they were right or wrong, fair or unfair.

What characterises this approach is the importance of identity, above all the identity of the individuals or group studied or studying itself [6]. Valuable work can be undertaken 'from the inside' or 'inside the whale': like Jonah in the whale, whatever the problems of objectivity and independence, there is a unique perspective and closeness to knowledge in the organisation. Identity forms a key part of sensemaking in organisations. We absorb and develop our identities, and our sense of identity, from interacting with others. Reflecting on events and states (the what happened and how we felt about it issues) allows us to further develop our sense of identity. Sensemaking needs a sensemaker [7].

Ethnographically, sensemaking leads to attempts to interpret, first by the sensemakers (like employees or managers), and second by any researcher involved in examining sensemaking by employees or managers. We find that sensemaking is grounded on a strong sense of identity. It shows in many ways – a manager tells us today about a situation yesterday where, perhaps, a customer was dissatisfied with a product and wanted it replaced [7, 8] Such managerial experiences are commonplace, generic and reiterative.

At the same time, paradoxically, they are unique and idiosyncratic. There may be a message there about how managers in that company should deal with all such customers in such situations. We also find a unique process of sensemaking going on, as the narrator identifies with the event while describing it. It is a story about her and the customer. It has a shape. It is in the past and so we have the perspective of the present. It may well take the form of a version which the narrator finds acceptable (for example, I dealt with the customer competently, or in such situations we are trained always to...). The matter of identity is important too because events come with states: how I felt then, how the customer reacted, how I feel now, and so on. One stage beyond that in turn is what the manager is prepared to disclose about the event and states of mind, and what persona they disclose to the researcher or observer.

Sensemaking, storytelling, and tacit knowledge

Sensemaking often takes us in an anthropological and ethnographic direction. It is heavily impregnated with self-referentiality, where people tell things from their own point of view, deliberately or unconsciously. Self-referentiality and the self-concept are central to the matter of identity, and, in situations where people in organisations can be persuaded to talk about events and states, self-referentiality and self-concept pervade the created texts. So socially situated are events and states, and so socially situated is disclosing them to others, that we may conclude, without stretching the point, that there is something about sensemaking which allies it closely to tacit knowledge.

If organisational knowledge is both explicit and tacit, and if tacit knowledge can become or be made explicit in deliberate ways, then the connection with ethnographic management research becomes more obvious. Explicit knowledge tends to be formally and normatively created and constructed. Substantively and procedurally, explicit knowledge may well have the formal hallmark of organisational approval – forms of thinking and ways of doing things which fit the canonical paradigms of organisational approval, into which new entrants are socialised, and through a thousand sanctions and incentives and behavioural cues experienced managers regard as second nature. This is the state of conscious competence. It is also the first stage along the spectrum of unconscious competence. In many organisations it can well be the state of unconscious incompetence.

Sensemaking comes before interpretation and explanation in organisations. In his classic study of sensemaking Weick [7] argues that sensemaking concentrates on how people generate what they interpret. Sensemaking is the process and interpretation the product. It involves constructing and filtering, framing and creating plausible ways of dealing with things, including the very awareness of subjectivity in the sensemaker. 'It is concerned with meaningful action, how individuals make sense of their situations and thus come to define and share realities that may become objectified in routinized ways.' [7, p. 15] Because it comes before interpretation and explanation, it often gets submerged into decision making: after the event, all most people remember are the decision and its consequences. Individuals, then, make efforts to understand and to find meaning in their experience, meaning to them, hence the centrality of identity in the process [8, 9].

The links sensemaking has with tacit knowledge appear in other ways. It is clear to see that managers talking about their experience will interlard the 'facts', as externally verifiable and publicly knowable, with personal impressions and inferences. This makes their account unique. It may not make it true, although, of course, it may be 'true' to them. It might even take the form of a conscious or unconscious deception. This is often the case when subordinates have to account for themselves to bosses, or where members of an organisation decide, through guile or fear, to purvey a PR spin on a story to keep the organisation clean in the public gaze.

Being 'tacit' in such situations, then, has individual or personal consequences in terms of the identity and self-concept of the narrator, and may well come about because they regard their personal identity as part of a corporate identity to which they believe they have to conform [10]. Many successful organisations socialise employees into defining themselves as members of an intellectual hive. Increasingly, meaning in social interaction is an inter-subjective matter, something where meaningfulness is a shared experience, a process where cues and symbols form part of a web of shared meanings. This can be seen pre-eminently in cyberspace in Usenet communities.

Being tacit also entails something else: accounts of events and states may be plausible rather than true. Since sensemaking, like organisational behaviour generally, arises in socially situated ways, it is enactive in consisting of things which are both created by the actors as well as described (usually later) by them, and takes place in group contexts where being socially acceptable is often more important for self-concept and personal well-being than being right, then it is no surprise that sensemaking has this strong epistemological dimension. Epistemology, of course, deals with the examination of what is true, what we mean by true. It has links here with what is right and fair, what we mean by right and fair (in other words, the ethical dimension). These things appearing so, it follows that sensemaking will not fall neatly in line with strict rationalistic assumptions about explanation (i.e. based on information, choices between known alternatives, rational choices followed by effective control and feedback).

The anthropological and ethnographic approaches have always acknowledged the importance of accommodating both objective and subjective elements. Being subjective is not necessarily pejorative, as 'scientific and experimental' methodologies would imply [11]. In the 'subjectivity' – assuming the evidence is reviewed clearly and gathered authentically and reliably – lies critical and elusive insights and revelations from actors directly implicated in the field of study. So being plausible rather than 'objectively' true characterises many ethnographic accounts. This is not to prevent managers or researchers or interlocutors seeking out other evidence which may corroborate or invalidate what is discovered. Indeed, *rashomon* (or diversified viewpoint research) can open up multiple perspectives on events and states, and is widely used in scenario planning [12, 13].

Dealing with such sensemaking encourages us to accommodate the many and often transient performance characteristics of ethnographic research. These take the form of smiles or gestures or mannerisms, which may, in body language or haptic terms, confirm or appear to confirm, contradict or appear to contradict, what is communicated as text or image. One of the recurring features of ethnographic approaches to managerial sensemaking concerns performance – the very telling of the story, the coming together at watering holes of communities of interest in the company, the gossip on the email, the self-dramatising behaviour associated with meetings and conflict and alliances in the workplace.

Any ethnographic research has to take account of the ways in which identity, plausibility and acceptability, and performance influence the character and mediation of the evidence. Again, observers and researchers (who may be other employees in the company attempting to get information about something, or external researchers or consultants trying to make coherent sense of the hows and whys of office or shop floor practices) may regard performance as a added value, providing an added insight into the meanings and interpretations. The account of the manager and the customer and the product that needs replacement may be mediated in self-ingratiating, self-deprecating, self-justifying ways. Laughs may hide chagrin. Formal gestures seek to convince that standard procedures were followed. The tacit becomes explicit.

Unconscious tacit knowledge, that far end of the spectrum of unconscious competence, is also detectable here. For anyone investigating knowledge management, and the exteriorisation and formalisation of tacit knowledge in the organisation, this domain is of particular value. Not only does it reveal, or appear to reveal, what people know, and know they know. It also reveals, if only inferentially, what people want others to believe they know, what people reveal unknowingly as not knowing, what they think of the company, and even what they really think.

Research often highlights meaningfulness as an essential ingredient in job satisfaction: it is also critical here where we are looking for that meaningfulness which is not necessarily the formally sanctioned meaningfulness of company training manuals and the many socialised public practices and rituals of the company. The meaningfulness here consists of what people see as meaningful, possibly what they are prepared to admit or imply as meaningful, and what we, as interlocutors and interpreters, might infer they believe as meaningful.

Plausibility and performance

Central to any anthropological study of organisations [9] is the view that truth comes from both the facts and people's beliefs and values. Leave one out and you get only an imperfect view of the truth. Knowledge management is a field where taxonomies of knowledge have shown a fixation with cognitivist paradigms. One reason lies in its pedigree in the world of information and computing. But that is no good reason not to deconstruct such paradigms so that they become more hospitable to the ethnographic insights, above all those associated with the indexicality of truth. This hinges on the ways in which apparently similar discursive tokens (like concepts, slogans, mission statements, staff manuals, and the rest) are taken to be publicly and consensually accepted as meaning the same to everyone. Indexicality suggests that connotation, situationality, subjectivity, and opportunity entail. Even these taken-for-granted or routinised meanings are themselves questionable.

Often the tangible vehicles through and by means of which meaning is expressed, like the information management systems and generally shared professional and company jargon and in-speak, represent such rich tacit knowledge only faintly. In sensemaking, which is preparatory to explanation, which is self-referential and identity confirming, with strong elements of plausibility and performance, it is no surprise to see how imperfectly current knowledge management performs in eliciting unconscious tacit knowledge.

Even here, however, a paradox hits us: that of the ethics, let alone the technological capability, of capturing unconscious tacit knowledge for company use, while at the same time guarding against imputations of surveillance and invasions of personal privacy. There are also issues of relevance: arguably, much unconscious tacit knowledge is unrecordable, transient, impressionistic, provoked by the emotion of a moment, flighty, created in order to impress or deceive, meaningless spin, or the party line The very process of elicitation, by exteriorising what hitherto was hidden, is vulnerable to manipulation.

On the other hand, it can be said that managers are known not just by their official pronouncements but also by their *obiter dicta*. These are transient remarks or asides, usually embedded in the normal ebb and flow of discourse, which nonetheless are revealing of the way the narrator really feels. So an assertion by a manager that his company had come out of due diligence with flying colours and, as a result, venture capital would be flowing in so that the initial public offering next month could be trusted by shareholders is fine *per se*, on the surface, for what it is. If the statement contains *obiter dicta* such as 'you may have heard otherwise but...', or 'John told me and he should know...', or 'that's why we're better than Oracle...' find their way into the statement, as they often do, then we have a level of explanation which reveals the sensemaking underlying it.

Such sensemaking shows that the narrator has been able to construct, filter and frame the explanation as a result of going through these stages. Or that he wants to give that impression. Or, most revealing of all (and far more commonplace than most narrators think), he does not know that he is giving that impression. This is why *obiter dicta* are critical cues for real meaning. Quite often narrators become aware they are using *obiter dicta*, and seek, by acknowledging their existence, to hide or neutralise them. So, the phrase 'you may have heard otherwise by...' is itself interpolated or glossed by a meta-phrase like 'you know what the financial journalists are like...', or 'people in your neck of the woods are always cynical...', or 'you probably don't know all the ins and outs of it...', or 'there's really no need to believe all you hear...'.

It is clear that unconscious tacit knowledge is being revealed here. It is also being revealed, or at least in sufficient ways for a vigilant interlocutor or colleague or researcher to pick it up. We can see sensemaking at work, self-referential, socially situated, inter-subjective. Where the narrative consists of formal explanations (like that manager is telling us his company is really really good for an investment), *obiter dicta* will frequently reveal the underlying sensemaking. Such *obiter dicta*, of course, may also take the form of performance cues, like a raised assertive voice, a doubtful tone, a careless laugh, a snide remark, a cynical aside. More important still is how, in this ethnographic elicitation of knowledge and meaning, we are finding an amalgam of facts and beliefs.

This characterises the ethnographic approach: that of capturing not just the events and states but also the actors' attributions of meaning to them. They are actors, part of the action, causative of action perhaps or victims or recipients of it. They are in it, part of it, authorities on it, so close at times that they can see it only one way, and, collectively speaking, very unlikely to have exactly the same view of things. Only when formal mechanisms of control, or the 20:20 wisdom of hindsight comes down like a shutter (a lot of storytelling, say in case studies and scenarios used in management training are retrospective), are the interlinear glosses managers add to the narratives removed.

So we get facts and beliefs. Facts given both denotative (objective meaning) and connotative (subjective or indexical meaning) status, beliefs and perceptions coming through, often obliquely, hinting at the existence of values beneath. Cognitive dissonance might show as a manager 'tells' one story (say, about how wonderful his company is) and 'shows' another (say, how he clearly feels he has to be one of the boys and follow the party line).

Tacit meta-knowledge

We get one more critical thing: an insight into the working of *tacit meta-knowledge*. The *obiter dicta* and performance cues reveal how much the surface script is very much the tip of a knowledge iceberg, mostly under the surface [4]. For many managers – bosses trying to get a clear idea of the progress of a project, employees working in teams sharing information but wondering how far they are really in the loop, consultants attempting to understand and improve processes by asking the people who run them, journalists getting doctors and policemen and politicians to explain their actions and policies, teachers evaluating exactly what students appear to know – the tacit knowledge is the most valuable kind of knowledge. It is the knowledge that really dishes the dirt, which really tells you what you want to know, which goes to the heart of the matter, which cuts to the chase.

That is why financial reports in the business world tend to cost so much – reports, say, on company performance or valuations or employee reward, by being difficult and expensive to gather and by being up-to-the-minute, cost a lot because managers are prepared to pay a lot for them. They are believed to be unique insights into the so-far hidden knowledge of the field: getting that knowledge helps you understand it in a deeper way than before, it gives you a competitive advantage, it allows the manager the 'aha!' configurational experiences which provide the best insights in creativity and innovation. In the same way, tacit knowledge has that added value. It is what we are not necessarily told, it is what we think but cannot confirm, and it is what others appear really to think about what they say to you.

The argument has drawn on *obiter dicta* so far, and they are revealing. In fact they indicate that the self-referentiality and performance elements of narrative are being used, consciously or unconsciously, as part of the explanation. They also indicate that the very act of mediating experience in narrative is itself part of sensemaking: it is socially situated and interactive each time it takes place, and such narratives are repeated, and adapted repeatedly. Narrators are aware their accounts change. They may change through circumstances, or because of power play between interlocutors, or because there is an audience of outsiders. The narrator is aware himself of himself as narrator, and that self-awareness leads to new textual explorations and performances. The script is always changing, growing, fading, new meanings being added, and new coherences being observed.

But it is not just the asides. It is the very structures of narrative that provide insights into tacit knowledge. And because storytelling is a discursive, flexible, protean, and idiosyncratic process, it allows for *tacit meta-knowledge* as well (in other words, where tacit knowledge is revealed and, as it is revealed, reveals itself to the narrator, who then incorporates this awareness into the very matter and manner of the storytelling).

Narratives take many familiar forms. We know this from folk tales and epics, which often follow prototypical forms like quests or contests or love triangles, and demonstrate beginnings, middles and ends, hinge on peripeteia or lead to catastrophe. Characters too often come in recognisable forms – the young hero, the black villain. We often represent our experiences using narratives – our own lives unfold chronologically, we make sense of a pattern of events at work chronologically, we look back and try to make sense of a sequence which has finished. We have expectations about resolutions and endings. We attribute good and bad, fair and just motives to the characters, and feel good or bad or sad or happy about what happens. Facts and feelings combine in complex ways. Stories exist for entertainment but also as part of sensemaking: they play roles as explanation, memory, and bardic interpretation. They act as folk memory, incorporate symbolism and history, and provide lessons for others and future guidance [14, 15, 16].

These general characteristics apply to organisations and management. Managers often represent their experiences in narrative form [17] It might refer to the strategic performance of the company over the last year, or the effect of restructuring the department because of introducing a network, or the way the company rewards and promotes certain types of staff, or how a particular manager coped with redundancy. All are strongly self-referential, not only in terms of the subject but also because managers are involved with the 'plot' of their narrative – as players and as tellers. Such narratives may represent unique or routinised situations. They may also be plausible rather than true, and represent an 'official' or biased viewpoint or agenda. In their individual and corporate ways, they enable the actors to make sense of what has happened and what is happening, and understand their impact and coherence better. They may also be managers whose role it is to ensure that others, like consultants, or journalists, or shareholders, understand better.

Sensemaking using narrative digs deeper than *obiter dicta* into the very structures used by narrative. These may be chronological, as the narrator looks back over her career, or over the last five years of the company, or what led to the takeover. Some narratives will include large components of financial information: it is possible to regard the annual report as a form of narrative. Yet chronology is just the start of this. Structure shows itself in deeper and ultimately more revealing forms, and these shed important light on knowledge in the company. A critical aspect of structural meaning in narrative lies with cause-and-effect.

Cause-and-effect

In real life things happen because other things cause them. Company aims predispose policy and planning in certain directions. Restructuring the firm entails new roles for some and no roles for others. Learning new information changes decision-making. Rivalry between sections may nurture high performance or breed stress and disillusion. Changes in the supply chain may result in cost savings. Cost savings may be sought and suppliers changed.

This cause-and-effect process also characterises stories [3, 15]. Many of the main narrative motifs hinge on a cause-and-effect logic, say that of a prince wanting to marry a princess, a hero going on a journey to the end of the world, a man drawing a sword from a stone and becoming king, a shoe fitting a foot enabling a young woman to become queen, a large wooden horse being gifted to an enemy city with fatal consequences. The very '... and then... and then...' sequence which characterises stories, particularly if they use suspense and surprise, typifies this. Consequentialism is an important ethical dimension of this, picking up on the need to consider decisions in terms of what their consequences or outcomes might be for those involved, and how right it is to make decisions on those grounds. Ends-means analysis, too, is a familiar managerial approach, in which the series of events leading up to and away from a decision need to be carefully analysed and mapped.

One of the most elusive features of best practice is to match the best decisions with the best outcomes. Areas like investment are difficult because of risks and uncertainty. A lot of managerial effort goes into peering into the future, say using narrative in scenario analysis [12, 13]. The explicit knowledge may allow or help managers to answer many routinised questions. It may help them to deal with important strategic and operational decisions – having accurate and timely information about sales and markets, design specifications and costs, are cases in point. Good managers, and especially outstanding ones, particularly the seers and gurus of management from Deming to Vassos, are attributed with having a configurational hold on the cause-effect relationships to be found in organisational environments. They see the connections between apparently disconnected things, notice the impact of independent variable x on dependent variable y, look holistically at different areas of company finance and make meaningful decisions

This is all explicit. But, iceberg-like, much tacit or implicit knowledge remains hidden away. The ethnographic approach again suggests a way through this difficulty, and again it is in the shape of narrative. It happens all the time in organisations, forming the fabric of work, often only partially noticed because people are busy, focus on their own tasks, do not always listen to each other in groups, ignore fools and knaves, pay studious attention to orders, resent some of them. Anecdotes go round about how Jane was promoted over Jim, consultants now get BMWs as perks, or 'the Glasgow office is being closed down'.

Narratives will have explicit structures. These will come through in many familiar ways – once upon a time to start the story off, 'and then' connectives to maintain suspense and act conjunctively to connect up parts of the story, and a build up to a climax like a final battle or death or reconciliation. Under such explicit structures there are implicit structures· implied connections between parts of the story, between mind-states of the characters and consequent actions. For example, 'the woman felt terror mount ... and reached for the revolver in the drawer...', or 'after the humiliation of the presentation, Sally vowed she would never let herself get into such a situation again....'.

Looking for tacit knowledge

Of crucial importance where implicit narrative structures are concerned are the *presumptive connections* made between the parts. These presume a logical connection, a coherent cause-and-effect sequence, and one which both narrator and audience can conveniently accept and understand. Such presumptive connections draw on *assumptive knowledge and beliefs*. These are to do with those assumptions people make, and have to make, in order to make the connections meaningfully coherent. So a narrator may say this: 'The sales figures have been good, but overhead rises are jeopardising our bottom line.' The assumptive knowledge is that the audience know the connections between sales and the bottom line (probably by way of sales and costs and gross profits, and then gross profits and overheads to net profits and the bottom line). Such knowledge may be explicit – after all it is known to most accountants. But in this situation it is also implicit – the presumption is made by the narrator that the audience will have, and use, its assumptive knowledge in order to make the connection and understand that part of the narrative.

Sensemaking is actively, and interactively, going on. Subtler and more elusive forms of implicit knowledge may be needed to make the connection. Narratives are often complex affairs even when textually short. A typical story might be one where a section has to close. The story might run:

> ... That March was rough for all of us. There'd been rumours for weeks, but when it came it hit George and the others really hard... [Um] Of course, the company as a whole had been doing quite well – well enough in fact to open another office in the south – somewhere Guildford way, I think – and that meant – top brass knew it all along – something would have to go – it's always the weakest link. Anyway, they'd done their best, and it was February, no January, when the figures simply got up and spoke for themselves – the local guys did their best, of course, but they had no clout. And so there it was – six guys on the scrap heap, including Mary just married. Very sad but I suppose we'd better get on with it...

The gist of this narrative is that a section of the company closed down. Telling the story is an active process of sensemaking, so we should not be surprised that it follows some kind of coherent narrative pattern – broadly chronological – and reflects a cause-and-effect logic. The chronological sequence is clearly from the end of the previous year up to the crisis in March. The narrator is looking back later that same year. The cause-and-effect logic is broadly clear: although the company had been doing well, the section had not, and finally poor figures did for them. The company did not help by opening another (perhaps rival) office elsewhere. This has a clear explicit logic which most managers reading or hearing it would understand. The assumptive knowledge there draws on our familiarity with the world of business and what happens there.

More elusive and implicit, more tacit, is the cause-and-effect logic underneath that. This draws on facts and interpretations of facts, some based on feelings. Already events have marched on, because the narrator is not exactly sure of dates. The narrator wants to give the impression that he saw it all coming, and that in a way the section could not have expected different. There are external business facts, but the narrator's cause-and-effect sequence loads it differently by suggesting that, if the company had wanted to save the section, it could have done so. Instead, it opened up a rival office. His sympathy for the section, and above all for some of the people in it, comes across: they did their best but ultimately it could not go on. Finally, he distances himself from it with the truism that life goes on: he will be able to live with the consequences, because ultimately it did not affect him directly.

In this way, we can begin to elicit the tacit meaning out of the narrative. We are getting an insight into the event and mind-states of the characters, including the narrator's own, in a way that may well never come through in any 'official' form. The narrator draws on his natural knowledge of the story form, as the audience does, in order to understand the meaning. But the narrative does not merely communicate the facts. It is mediated through a complex and idiosyncratic filter of the narrator's consciousness, what he says and what he agrees to share with the interlocutor.

Moreover, the cause-and-effect logic of the story is as important as the 'factuality' of the story which is where it all starts. The narrator's own interpretation of the way in which one thing caused another, the actions or inactions of particular agents in the story had impacts on particular recipients, make presumptions about what we, jointly and intersubjectively, regard as coherent and relevant. We can then begin to start examining the domain of assumptive knowledge there. This deals with why the narrator and the audience (perhaps we ourselves) assume a connection between things, and the extent to which we accept a causative link between them. Some events may, after all, be or seem so remote from one another as to make causation implausible.

Part of the assumptive knowledge domain deals with the assumptions themselves and how assumptions are formed. Experience external to this event is part of this, but, looking just at the situation, we are drawn to make certain assumptions or warranties about the connections. That certain things happen, in business generally, and in this situation in particular, because other things happen. Because top management neglected the section, it got closed down. Because the section did not seem to have enough autonomy to act effectively, it failed. Probably both happened at the same time, making closure unavoidable. The warranty is that logical connective which enables us to argue for and then accept the cause-and-effect sequence.

Role of narrative knowledge

We might then go on to generalise from the narrative. This is one of the most frequent uses of narrative in organisations. They are unique and idiosyncratic, but they are used as a key part of sensemaking because they tell us something important about the circumstances now, and they can also be used as pointers or lessons for the future (e.g. if the company were to do this, then similar or comparable outcomes would occur). They can also be generalised from and add to the store of exemplars of successful and unsuccessful management practice which becomes the folklore of management itself

The implicit knowledge here, revealed by *obiter dicta* and cause-and-effect analysis, is, of course, tacit knowledge. In being expressed and shared in narrative, however confidential the setting is, it can be defined as conscious tacit knowledge. So the last step is to look for the unconscious tacit knowledge, that meta-knowledge which we have argued is so important for knowledge management.

The narrative provided above is full of unconscious tacit knowledge clues. Part of it lies in wisdom attributable to the general human condition – the fact that the closure hit the section really hard (there is always a weak link), the fact that figures can get up and speak for themselves, and that when things like this happen people end up on the scrap heap of life. Part of it, too, lies in the tone (the performative dimension of narration), saying that the company had been doing well, well enough to open another office. The view that top brass knew it all along and probably had an agenda all along, hostile to the section and in favour of Guildford. And finally, the very self-referential (even self-serving) comment that, even though it is very sad, life goes on.

We might infer much about the unconscious tacit knowledge held by the narrator. One thing for sure is the gestural nature of the remark about it being sad. He does not want us to think he is unfeeling. He ducks his head below the parapet and takes a pro-company position at the end. Yet the tone up to then has implied heartlessness and blame on the part of senior managers, a them-us dialectic which the narrator uses but finally dissociates himself from. So up to a point, then, the narrator demonstrates an ability to reflect on his own experience, including the very experience he is using to represent his experience, the narrative itself.

What he knows he knows, in and from the story, is conscious tacit knowledge. What he knows he knows as well as what he, unconsciously, reveals to us as knowing, forms the unconscious tacit knowledge. We are assuming that he is in a position to know. He may be pretending or deceiving us. If we are insiders, then we also have knowledge, but knowing that his method of telling us may change. He will know that we know, we will know that he knows that we know, and so on. The effects of this, in eliciting reliable evidence, casually or formally in research, must be considered carefully.

The other revealing area of tacit meta-knowledge is that of feelings, beliefs and values [18]. Since sensemaking is socially situated, and since most discourse is subjectively determined, it is important to take account of this The narrator is indignant about the closure. Or so it appears. Or so he wants us to think. He wants us to think him fair-minded, not gullible in believing that senior managers really care about their employees, even if they do. He seems to have prejudice about opening offices in areas where employment is said to be good. He thinks it is all very sad, although the real meaning seems, indexically, to be 'inevitable'.

Knowledge management emphasises the importance of identifying, storing, and using the whole knowledge and creative resource and potential in the company [19]. The stark dichotomy between explicit and tacit, or as Polanyi says between public and private, is fine as far as it goes. Attempts to build up corporate memory and capture professional decision making expertise and heuristics in software programs succeed only up to a point.

If knowledge management is truly to capture knowledge – and there are ethical reasons why this may not be a good idea – its search for tacit knowledge has to extend further than conscious tacit knowledge to unconscious tacit knowledge. That is, in a sense, an area we might call wisdom management, or prejudice management, or ignorance management – for all these characterise the complex, usually hidden, area we have been exploring An assumption made here, too, is that employees will willingly and fully co-operate in programmes that elicit wisdom from them.

For many knowledge workers, there will only be a fair exchange if their own learning is at least as substantial as anything they pass on to others In a world of transactional employment contracts and portfolio employment, many employees might convince themselves that becoming part of a corporate memory has few advantages and many disadvantages for them. Knowing that, and knowing that you know that, as well as the innate self-referentiality of human beings, may lead to fewer employees taking a fully authentic and sincere part in knowledge management than management would like to think.

References

1. Davenport, T H. and Prusak, L. *Working knowledge. how organizations manage what they know*. Cambridge, MA: Harvard Business School Press, 1998.

2. Nonaka, I. and Takeuchi, H. *The knowledge-creating company*. New York and London: Oxford University Press, 1995.

3. Hewstone, M. *Causal attribution· from cognitive processes to collective beliefs*. Oxford and Cambridge MA: Blackwell, 1989.

4. Schank, R.C. *Tell me a story: a new look at real and artificial memory*. New York: Charles Scribners' Sons, 1990.

5. Deal, T.E. and Kennedy, A.A. *Corporate cultures: the rites and rituals of corporate life*. New York: Addison-Wesley, 1982.

6. Van Maanen, J. *Tales of the field· on writing ethnography*. Chicago and London: University of Chicago Press, 1988.

7. Weick, K. *Sensemaking in organizations*. Thousand Oaks, CA, and London. Sage Publications, 1995.

8. Polanyi, M *Personal knowledge*. London: Routledge, 1958.

9. Wright, S., ed. *Anthropology of organizations*. London and New York: Routledge, 1994.

10 Martin, J., Feldman, M.S., Hatch, M.J. and Sitkin, S.B. The uniqueness paradox of organizational stories. *Administrative Science Quarterly*, 28, 1983, 438–53.

11. Ellis, C. and Flaherty, M., eds. *Investigating subjectivity: research on lived experience*. Newbury Park, CA and London: Sage Publications, 1992.

12 Schwarz, P. *The art of the long view*. New York: Doubleday, 1991.

13. Van der Heijden, K. *Scenarios: the art of strategic conversation*. Chichester: Wiley, 1996.

14 Brawer, R.A. *Fictions of business: insights on management from great literature*. New York and London: John Wiley, 1999.

15 Paulos, J.A. *Once upon a number: the hidden mathematical logic of stories*. London: The Penguin Press, 1998.

16. Propp, V. *Morphology of the folktale*. Austen, TE· University of Texas Press, 1968.

17. Snowden, D. Story telling: an old skill in a new context. *Business Information Review*, 16(1), 1999, 30–37.

18. Augé, M. *Non-places: introduction to an anthropology of supermodernity*. London and New York: Verso, 1998.

19. Stewart, T.A. *Intellectual capital: the wealth of organizations*. New York: Currency/Doubleday, 1997.

A future for publishers serving technology industry markets: adapting to the opportunity of the Internet era

Paul M. Evans
Elsevier Advanced Technology, The Boulevard, Langford Lane, Kidlington, Oxford
p.evans@elsevier.co.uk

The future possible roles of the publisher in industry markets are considered, especially for firms close to high technology. It is forecast that the Internet will become the dominant business space for firms. Even in physical manufacturing the advances of the Internet are becoming important with changes in supply and distribution patterns. In order to meet the new era of opportunity three scenarios are presented. Firstly, continuation of the current one with the publisher as *information provider*. Secondly, the publisher may become an *orchestrator of interactivity*. Hence they can be the hub of a new intelligence facility for firms to know their external environment better. This also would be an appropriate response to the expected steep increase in requests for more customised sources of information. Lastly, the publisher can go a stage further and become a *consultant* or facilitator of the agile enterprise based on the Internet. This would enable the client company to become a more adaptive and proactive entity afterwards. Recommendations are made for assisting the shift for these further scenarios to become possible.

Introduction

The rise of the so-called new economy is viewed with caution as technology stocks peak and trough and pundits warn of the likely shakeout in the sector of Internet companies. It is known that for many such companies a key strategy has been to attract investment capital and for the entrepreneurs to take their rich pickings early, rather than betting that each company will genuinely make it to becoming a profitable enterprise. Despite these concerns, it is likely that the new business era will come into being from this foundation on those Internet businesses that prove to be successful in meeting genuine customer needs as suggested by Hagel [1].

For publishers this at once represents a threat of new competition, an opportunity for repositioning their businesses to meet the new economy and brings the prospect of higher margins than the average in the traditional physical environment. For publishers serving industrial companies, especially in high technology areas, the opportunity is doubly relevant as the market itself is assessing the same opportunity to take advantage of doing business in the Internet economy. In serving the needs of the industrial markets these publishers need to take full account of the extent of the change that is likely to happen to their customers. It is not simply a question of firms changing their product lines. They will be involved in radical restructuring so that the very basis of the firms will change.

The change publishers foresee includes the increased need for more customised sources of information. Also it is clear that needs may change as the Internet economy moves from an economy based around manufacturing to one which is more service-oriented.

The time is ripe for this change due to the proliferation of the computer since the arrival of personal, and now, more significantly, networked computing to businesses and into homes. This has led to a move from transactional computing (such as in order processing) to the manipulation and presentation of information itself, giving rise to a revaluation of information as a commodity. Now, even beyond that, is the revaluation of communications and the ability to:

1) establish stronger relations and faster interchanges to increase the pace of new product and service developments coming to market;
2) enable firms to restructure relationships with other firms through alliances for complementary offerings (the era of 'co-opetition') and to outsource activities where adding value is not maximal;
3) establish an economy where traditional physical values (as originally represented in the supremacy of manufacturing in successful national economies) are being superseded by a 'dematerialised' economy with increasing importance placed on developing knowledge and know-how.

Essentially this results in greater emphasis on the service economy. With more services there is a further increase in knowledge intensive activities in the industrial sector and the services provided are requiring that firms have more and better information and communication with both their suppliers and their customers. Industry is requiring more information to do its job properly and the general level of knowledge intensity in the economy is changing rapidly.

The increasing stress on connectivity gives rise to industry players needing to be part of networks around their business and the need to gain access to these networks is critical to future business success. The time of the network has very much arrived and the age of the PC is over. The value of business can be increased, but for publishers the task is to meet this need for value with enough relevant content. The customers of the publishers in these markets are looking to see what packages of information can be provided from information on processes and innovation close to science and technology, to information on trends of customers and trends within the industry as a whole (including competitor information).

Environmental concerns will also be increasingly important guiding principles for economies due to social and political pressure [2]. Environmental engineering has shifted from solving the 'end-of-pipe' problems of pollution in manufacturing processes, to a concern for life cycle analysis of physical products, as Braun and Elliott [3] note. So, for a washing machine the cost in environmental terms, including energy expenditure, is looked at not just in terms of producing the product, but in the longevity of the product as well as the cost in the use of the product over its lifetime (and also factors in its final disposal or recycling). This ties in with the increased emphasis on the service economy that accompanies the Internet era. Just as there has been an increase in the number of cars leased rather than purchased, so in many current areas of manufacturing the move to the service economy will be based on customer needs being met, rather than in shifting more products. This will be just one of the new areas of topicality that customers in industry will be looking to publishers for in being able to monitor their external environment.

In a more networked economy, the premium on relationships with existing employees, customers, suppliers and partners (across all parts of the value chain) is being increasingly unlocked via new communication links. Therefore, publishers serving these new types of organisation will have to be relevant to this transition in their customers' businesses from current models in the physical manufacturing economy. Many will establish not just shopfronts on the Internet. They will shift their entire processes, via technology, into what is almost a new dimension as the service economy increasingly predominates.

However, despite being faced with these changes, publishers have essentially been a relatively conservative industry [4] in terms of technology and development of best practice (such as total quality management (TQM), or total productivity management). TQM and related practices were in contrast taken up with enthusiasm by many of these publishers' customers in other industries. Some of those publishers closest to the technology industry have been rooted in the relatively profitable STM (scientific, technical and medical) sector and the drive for getting the greatest efficiency out of these publishing firms has, to some extent, been inhibited, ironically, by this record of economic success. So there is a risk of too much inertia, even while it is clear that many publishers are adapting to automating their product lines via the Internet. Some are going further and inventing new product and service concepts to take further advantage of the Internet opportunity. But, whether the large existing traditional publishers or the smaller Internet start-ups provide the basis for the new business (probably a remix of both is more likely), it is unlikely that merely acting as information providers or content brokers will meet the needs of these new organisations. Something more radical needs to be considered and this will of necessity change the role of the publisher in this sector and require a more intimate understanding of content and processes relevant to customers in industry.

Scenarios for the future for publishers in technology industry markets

Three scenarios can be painted as to where publishers may take their business to meet the new needs arising out of the major economic changes in their customers' businesses. These scenarios are conceived in relation to a model developed by Papows [5] in relating data, information, knowledge and work to: a) the empowered individual; b) the automated workgroup (which may itself sound like a rather dated concept in the network age, but which suits Papows well as the chief executive of Lotus which provided software for such groups), c) the integrated enterprise; and d) the extended enterprise. These can be related to the ways in which the customer is organised and how the publisher can support its customers' activities.

In painting these scenarios against Papows' model it is vital first to understand how Papows believes that firms work.

Papows' model as an analytical context

In new organisational designs, particularly with knowledge based networking organisations coming into being, it is possible to see some of the elements that had predominated in earlier designs coming back in different ways. Papows' model draws on the key organisational models of the past but places them in a new context. As an analytical model it is useful, even if in reality organisations are more complex and really represent combinations of different types of organisation operating simultaneously. This is part of the complexity that publishers must come to understand if they are really to better meet the needs of their industrial customers Only through understanding their processes and organisation can they hope to move into the higher premium value activities beyond merely importing information to the doorstep of firms. As an analytical framework it gives insight into where the publisher can focus attention for adding value to the customer's business in the mid-term and into the longer term.

In the model, data refers to unqualified facts. Information refers to data with context Knowledge is information in terms of relevant know-how or know-why. Work is the combination of data, information and knowledge in the creation of action in the business.

Meanwhile the types of organisation to consider are the empowered individual capable of steering his or her own activity; the automated workgroup (which brings to mind the autonomous workgroup of the late 1980s and how this gave rise to quality teams); the integrated enterprise which emphasises the increased reliance on corporate systems; and then finally the extended enterprise which reaches into suppliers and ultimately through to customers with the Internet. These can be represented following Papows' model in Table 1.

This model is a useful reference in the definitions of the three scenarios that follow and the shading indicates how the different areas of Papows' model relate to the options available to publishers in the industrial market.

a) The publisher as information provider ('at the doorstep')

In being an information provider (more or less as at present) the publisher can allow access to information for empowered individuals (usually subscribers) or gatekeepers. This relates to Leonard-Barton [6] and her dimensions for a core capability detailing how knowledge is generated and used. In her model (see Figure 1) this puts the publisher's role clearly in the area

Table 1. *The enterprise solutions framework (based on Papows' model) and its relevance to publishers*

	Data	**Information**	**Knowledge**	**Work**
Extended enterprise	Online sales and other transactions with customers and suppliers	Connecting to outside stakeholders	Operating alliances; working with/within market interest groups	Using cyberspace as the principal business space
Integrated enterprise	Building corporate databases and other corporate applications	Encouraging cross-functional communications	Leveraging intellectual capital and best practices	Re-engineering business processes
Automated workgroup	Using workgroup data systems and databases	Encouraging cross-functional communications	Enabling collective discoveries and decision making	Improving conduct and control of workflow
Empowered individual	Enabling data creation, access or usage	Enabling information access and authoring	Enabling creation, access and distribution of expertise	Assuring integration into workflow systems

Publisher as information provider

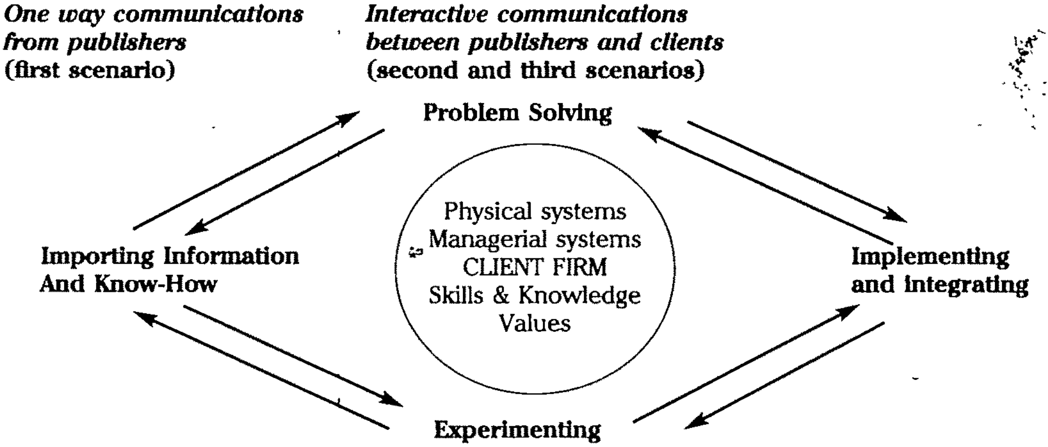
Publisher as orchestrator of interactivity

Publisher as consultant or change agent

of enabling knowledge to be imported. However, it has little relationship with problem solving, implementing and integrating knowledge or experimenting as discussed in a recent paper on knowledge synthesis [7]. It may impact on skills and knowledge, even to some extent on values, but is not related so closely to the managerial systems and physical systems in terms of a more intimate relationship.

Comparing this role of the publisher in both Papows' model [5] of the enterprise and Leonard-Barton's model [6] centred on knowledge activities, it is clear that the publisher is able to deliver a service that is required but does not enter far into the workings of the firm. There could, however, be further sharing of the information both within and beyond the firm with electronic publish-

Figure 1. *Dimensions of a core capability (after Leonard-Barton [6]) related to the prospective role of publishers in the Internet era in assisting the knowledge cycle of client firms*


One way communications from publishers (first scenario)
Interactive communications between publishers and clients (second and third scenarios)
Problem Solving
Physical systems
Managerial systems
CLIENT FIRM
Skills & Knowledge
Values
Importing Information And Know-How
Implementing and Integrating
Experimenting

ing technology, and content could be circulated by readers to other interested parties if they were to act as superdistributors of nuggets of information. This is a model that has been proposed for micropayments systems, with fees going back to publishers and authors each time the information is accessed within a market. It is a mechanism, which can therefore help in supporting workgroups such as research or product development teams, but it is unlikely to knit far into the fabric of the firm in terms of intranets etc.

In this role of the publisher there can also be a supply of content, hard information or data about which the publisher does not necessarily need to have a highly developed understanding. The publisher acts more or less now as a broker, e.g. of papers and abstracts and some digested material using expert editorial resource. For the STM publisher the key competence is in the quality management, especially for ensuring the review processes of primary information. However the key tasks are outsourced, such as abstracting and refereeing. In the more scientific areas the distribution guarantee is the critical competence.

However, the need in industrial firms is much more for digested information, for relevant content rather than necessarily the best science, so without the focus on value-added processes using content-close editorial skills, the offering is relatively limited to industrial firms. Even with the editorial added value which converts data to information by providing more context and relevance, more opinion and analysis, the role is still one of a delivery merchant leaving materials at the doorstep of the firm.

In this role the publisher is most likely to act as a stand-alone company rather than in alliances, although there may be linkages with other publishers fulfilling a similar role and joining together in delivering more comprehensive offerings.

b) The publisher as orchestrator of interactivity

It is perhaps not surprising that the full potential of the Internet has been slow to be realised. Interactivity is the real development which can take the publishing industry beyond the gains of mass production that came with Gutenberg and the movable type of the early printing presses. However, speed is still a problem. WWW is still humorously referred to as the World Wide Wait. Publishers who have set up chatrooms in professional settings have generally been disappointed due to the lack of usage. One way through this is clearly to speed up the Web. Another is to make more use of the editorial skills to develop more effective moderators of content. This requires a really intimate understanding of users and their problems in industry.

Information is not just provided via moderators and through the use of enquiry analysts or consultants, but cross-functional communications are also enhanced to allow notes features for people to add comments to information received. Online help facilities can allow for dialogue with information consultants or librarians as well as experts in particular fields. The publisher can help support a network of experts for its client firms and may allow information handling by the experts to provide leads for the same acting as external consultants. The publisher acts as a bridgehead to bring the problems of companies to the experts who can serve them. These facilities can be integrated into firms' intranets provided the publisher can manage issues of confidentiality of firms' information (even, for example, confidentiality issues about which areas merit investigation).

In this model the publisher develops as a moderator for the external environment providing inputs about political, economic, social and technological changes around the firm. As such it represents an interactive addition to the traditional Executive Information System which tended in the past to concentrate on internal company data. What is more it ensures the information flow is not just to an executive war-room, but to the key teams and functions within the enterprise.

The publisher is thus better able to support the workgroups and is beginning to support the enterprise more fully, while dealing not just in information or data but in knowledge since expertise is being imported. It is also important for the information to be as relevant as possible and the offering more customised.

For this to be effective there needs to be more of a culture of learning (the learning organisation), and an emphasis on the customer firm's need to get closer to its environment and to adapt to meet new competitive factors. The publisher is effectively acting as an intelligence operation for the customer and the role of orchestrator for interactive processes enhances the original role of information provider.

In this role there can be useful alliances made with Internet providers, or even software consultancies, or systems houses serving general information needs within firms.

c) The publisher as consultant or change agent

In the third scenario the publisher continues to act as an information provider and an orchestrator of interactive processes in order to provide outsourced intelligence operations for firms. More than this, however, it acts as a consultant or facilitator to enable the customer firm to manage the change to establish itself with the Internet as its principle business space. This is the top right corner of the diagram in Table 1 and impacts on the full-extended enterprise. The consultants or facilitators within the publishing or consulting organisation activate the move to establish alliances and to outsource areas of business where obtaining a value-added premium is not maximal. This also requires facilitation in there being enhanced re-engineering of the business processes through the application of computer technology for the Internet business environment. The technology is used to fix some of the intellectual capital and to proceduralize new relations into the business processes. However, this is not a one-off conversion of processes. It must result in the establishment of a new more adaptive business paradigm so that the company can react quickly to changes in the business environment. This is a stage beyond partnering the client firm to become a learning organisation. It is about partnering the firm to become an agile enterprise. In doing so the consultant works in conjunction with the external antennae of the company served by the publisher as an outsourced intelligence operation and also works closely with the areas of highest expertise and business understanding/decision making within the customer firm.

In this role the publisher has thus shifted from being a supplier to becoming a change agent and strategy developer/facilitator. Whereas in the previous scenario the publisher was shifting to act as knowledge engineer the publisher now moves into what Papows [5] conceives 'work' to be: i.e. the interaction of information, data and knowledge to create action. The new area of activity is as both knowledge engineer and work engineer with clear involvement in the processes to set up Web relations with the market, within the firm, across to alliance partners and suppliers. More than this, as a facilitator, the publisher or consultancy (for such it has become) partners the customer organisation in transforming its business and enabling itself to continue transforming into the future. In this role the publisher may choose to make alliances with major consultancy organisations or ones specialising in technology.

This option is a considerable extension to existing publisher roles and for large companies it would require considerable change. However, within these larger companies there are examples of smaller more agile units in specific industrial markets where this kind of development could be started and as implementations are rolled out to new areas the new role can be fulfilled. For the larger publisher merely to join up in broad alliances with large technology consultancies without this pragmatic base in specific vertical markets the exercise would probably prove futile. However, if developed niche by niche it could see the publisher stepping well beyond the traditional role, so characteristic of the STM publisher industry, of acting merely as a broker of information one step removed from the end-user market servicing libraries.

Implications of the various scenarios in terms of business models

Many of the implications have been described in the previous section; however, it is worth comparing the three scenarios in more detail. Especially, it is worth examining business models. (See Table 2)

For the information provider the key business models are based on subscriptions (for access rights or according to usage) and advertising revenues where incidental suppliers to firms can be enlisted in support due to benefits for their marketing communications.

For the orchestrator of interactivity and provider of an outsourced intelligence service the model can be based on a higher subscription or membership, while there is still the opportunity for a portion of supplier sales to firms if the publisher acts as 'infomediary' at the hub of business relations for the firm [8].

For the consultant or facilitator it is more likely that the revenues would be as in the previous example, but with fees related to the increased earnings of the customer firm as it succeeds in the Internet environment through more efficient operations and through increased business.

It is likely that the last scenario would lead to the highest margins as the value of the information within the business is more integrated and the relationship between the customer and the publisher/consultant more intimate.

Conclusions and recommendations

For large existing STM publishers, on the boundaries of providing information to high-tech industry, the task is to establish an enterprise-wide offering rather than merely serving empowered individuals or libraries/corporate information centres. In order to do this the offering has to be geared towards the customized intelligence service with the specific expertise of analysts and editor/moderators. However, to be one move ahead, this should be accompanied by an experiment with alliances with technology consultancies active in developing e-business for their clients. These consultancies will need to develop a fuller understanding of the role of importing knowledge of the external environment and digesting material to ensure relevance.

For smaller Internet start-ups there is value in working with consultants in niche markets around specific industries, while making use of other information sources tc supplement own-originated content b means of alliances with other publishing c Internet firms. If the model brings the success that is hoped for the customer firms then there is the opportunity to grow organically, or to go to consultancies, or larger publishers and work with them to extend into other

Table 2. *The three scenarios – in a table of comparison*

	Information product or service on offer	**How?**	**With whom?**
Publisher as information provider	Industry information news, technology process; business information	Supply of information to the doorstep of the company served	Alone or with other publishers on collaborative projects or linked systems
Publisher as orchestrator of interactivity	Customised information of specific relevance to the firm; answers to specific enquiries; access to network of experts	Integrated into Intranet systems and internal information systems monitoring external environment	With Internet companies or systems houses serving firms' information systems needs
Publisher as consultant or change agent	Re-engineering of business processes to replace current business models with Internet centred models in virtual space	Integrated services including consultancy offering to change way of working and capacity for adaptability	With high quality technology consultancies

ustry areas. Since the pace of change is ly to be fast there is a virtue in proving the del first in one area, but then quickly celerating the roll-out to other areas by rking with larger players.

eferences

. Hagel, J. and Armstrong, A.G. *Net gain.* oston: Harvard Business School Press, 1997.

2. Freeman, C. *A green techno-economic paradigm for the world economy.* Chapter 10. The Economics of Hope. London, UK: Cassell, 1992.

3. Braun, E. and Elliott, D. *Block 6: global patterns in technological innovation. (Course T302: Innovation: design environment and strategy).* Milton Keynes, UK: The Open University, 1996.

4. Charkin, R. A new paradigm for STM publishing. In: Blunden, B.M., ed. *Electronic Publishing Strategies.* Leatherhead, Surrey, UK: PIRA, 1997.

5. Papows J. *Enterprise.com.* London, UK: Nicholas Brealey Publishing, 1999.

6. Leonard-Barton, D. *Wellsprings of knowledge.* Boston, Massachusetts: The Harvard Business School Press, 1995.

7. Evans, P.M. Facilitating scientific communication for industrial innovation. *Aslib Proceedings*, 52(4), 2000, 150–160.

8. Hagel, J. and Singer, M. *Net worth.* Boston: Harvard Business School Press, 1999.